

Ramirez, Delia

From: Wesner-Early, Caryn (ASRC)
Sent: Tuesday, February 12, 2002 11:53 AM
To: Ramirez, Delia
Subject: FW: dates for 2 GenBank entries (allen) (UPD#43785)

Examiner Ramirez -

Here are your publication dates - let me know if you need anything else.

Caryn
Caryn S. Wesner-Early, MSLS
Technical Information Specialist
Biotechnology and Chemical Library
U.S. Patent and Trademark Office
Phone: (703) 308-4501
Fax: (703) 308-4496
caryn.wesner@uspto.gov

-----Original Message-----

From: update@ebi.ac.uk [mailto:update@ebi.ac.uk]
Sent: Tuesday, February 12, 2002 9:50 AM
To: Caryn.Wesner@uspto.gov
Subject: Re: dates for 2 GenBank entries (allen) (UPD#43785)

Dear colleague

You have requested release date information for EMBL entry(s):Z49617 and Z49260.

The date given in the first DT line of this entry is the date the entry first became available for public disclosure.

Example:

DT dd-month-year (Rel. nn, Created)

The date in the Submission Reference RL line

Example:

RL Submitted (dd-month-year) to the EMBL/GenBank/DDBJ databases.
is the date that EMBL received the submission from the submitter.

DISCLAIMER: PLEASE NOTE

No guarantee is given that release date given in the entry is accurate. The actual release date will depend on the availability of network services.

No guarantee is given that the entry is complete and accurate. In particular the sequence data of the entry may not conform to the sequence data of the original publication where the sequence was first disclosed to the public.

I enclose a copy of the entry(s).

Yours sincerely

Yvonne Allen

ID SCYJR117W standard; DNA; FUN; 1825 BP.
XX
AC Z49617; Y13136;
XX

THIS PAGE BLANK (0.000)


```

SV      Z49617.1
XX
DT      06-OCT-1995 (Rel. 45, Created)
DT      11-AUG-1997 (Rel. 52, Last updated, Version 2)
XX
DE      S.cerevisiae chromosome X reading frame ORF YJR117w
XX
KW
XX
OS      Saccharomyces cerevisiae (baker's yeast)
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
XX
RN      [1]
RP      1-1825
RA      Rose M., Koetter P., Entian K.D.;
RT      ;
RL      Unpublished.
XX
RN      [2]
RP      1-1825
RA      MIPS;
RT      ;
RL      Submitted (25-SEP-1995) to the EMBL/GenBank/DDBJ databases.
RL      Data collected by MIPS on behalf of the European yeast chromosome X
RL      sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am
RL      Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.embnnet.org
XX
DR      SGD; S0003878; YJR117W.
DR      SWISS-PROT; P47154; ST24_YEAST.
XX
FH      Key                Location/Qualifiers
FH
FT      CDS                343..1704
FT                        /db_xref="SGD:S0003878"
FT                        /db_xref="SWISS-PROT:P47154"
FT                        /note="ORF YJR117w"
FT                        /gene="STE24"
FT                        /protein_id="CAA89647.1"
FT                        /translation="MFDLKTILDHPNIPWKLIISGFSIAQFSFESYLTYRQYQKLSETK
FT      LPPVLEDEIDDETfHKSRNYSRAKAKFSIFGDVYNLAQKLVFIFYDLFPKIWHMAVSLL
FT      NAVLPVRFHMVSTVAQSLCFLGLLSSLSTLVDLPLSYSHFVLEEKFGFNKLTQVLWIT
FT      DMIKSLTLAYAIGGPILYLFLKIFDKFPTDFLWYIMVFLFVVQILAMTIIPVFIMPMFN
FT      KFTPLEDGELKKSIESLADRVGFPLDKIFVIDGSKRSSHSNAYFTGLPFTSKRIVLFD
FT      LVNSNSTDEITAVLAHEIGHWQKNHIVNMVIFSQ LHTFLIFSLFTSIYRNTSFYNTFGF
FT      FLEKSTGSFVDPVITKEFPPIIIGFMLFNDLLTPLECAMQFVMSLISRTHEYQADAYAKK
FT      LGYKQNLCLRALIDLQIKNLSTMNVDPLYSSYHYSHPTLAERLTALDYVSEKKKN"
FT      source              1..1825
FT                        /chromosome="X"
FT                        /db_xref="taxon:4932"
FT                        /organism="Saccharomyces cerevisiae"
XX
SQ      Sequence 1825 BP; 570 A; 397 C; 280 G; 578 T; 0 other;
acctaccttt ttttctatct tcaacaacga aacgccttac acacacacac acatacatct      60
acatacatat atacaaatat acatatatgt aaacttgtat attcattcct attaaccaaa      120
aagaggcaat taaacttttc cctctttttc tacgtcattt actcaaaaac tctaattcct      180
tcgtctctgt tctgccattt tctccagaaa aaaatcgacg ggaaataaaa aaaaaaagac      240
aacgaacaag agaaaaagtt cggaatttat aaaccacttc tataattaac aggaaaagga      300
aggaaaaaaa aggaggaaat agaaaactgc aggcctttat tcatgtttga tcttaagacg      360
attctcgacc atcctaatat cccgtggaaa ttaatcattt ctgggttctc gattgcccac      420
ttttctttcg aatcttactt gacgtacaga cagtaaccaga agctatctga aacaaagttg      480
ccacctgtgc tggaagacga aattgatgat gaaacttttc ataaatcaag gaactactcc      540
cgggccaaag ccaagttctc cattttcggt gacgtctata acctagccca aaagctagtt      600
ttcatcaaat acgacctctt ccctaaaate tggcacatgg ccgtttcttt attgaatgca      660
gtcctgccag tcagatttca tatggtctcc actgtcgcac agagtttatg cttcttggtg      720
ctcttatcca gtttgtctac cttggttgat ttgccactct cttactatag ccattttgtc      780
ctggaagaaa aatttggttt caataaattg accgtccaac tatggatcac cgatatgatc      840

```

THIS PAGE BLANK (0.370)

```

aagagtctga ctttggcgta tgctattggt ggcccaatcc tttacctggt ccttaagatc      900
tttgataaat tccctactga tttcctttgg tacattatgg tcttcttggt cgttgtccaa      960
atcttagcca tgacaatcat tccagtcttc atcatgccca tgtttaataa gttcactcca     1020
ttggaggacg gtgaactgaa aaaatctatt gaaagtttgg ccgatagagt tgggttccct     1080
ctagataaga tttttgtcat tgacggctca aaaagatctt ctcatc aaa cgcataatttc     1140
acaggtttgc cattcacctc caagagaatt gttttgttcg acactttagt gaacagtaat     1200
tctactgatg aaattacggc tgttttggcc catgaaatcg gtcactggca aaaaaaccac     1260
atcgtttaata tggtcattct tagtcaattg cacaccttcc tcattttctc ccttttcacc     1320
agcatctaca gaaatacatc attttacaac accttcgggt ttttcttaga gaagtccact     1380
ggcagttttg ttgatcccggt tatcactaag gaattcccca ttatcattgg atttatgtta     1440
tttaacgact tattaactcc actcgaatgt gccatgcaat tcgtgatgag tttaatttcc     1500
agaactcatg aatatcaagc tgatgcttat gctaaaaaat tgggctacaa gcaaaatcta     1560
tgtagggtc taattgatct acaaatcaaa aacctttcca ccatgaatgt agatcctctg     1620
tattctagct atcattattc ccatccaact ctagctgaaa gattgaccgc tctagactat     1680
gttagtgaaa agaagaaaaa ctaatctata gagtacacat attagcatgt accgttaaat     1740
tcagcttcgt tatgtctata tctacataca tacacaggta tctactataa gaataaagga     1800
aagaaaaaat aaacgattaa acatt                                           1825

```

//

```

ID   SC8156      standard; DNA; FUN; 29987 BP.
XX
AC   Z49260; Z71257;
XX
SV   Z49260.1
XX
DT   16-MAY-1995 (Rel. 43, Created)
DT   11-AUG-1997 (Rel. 52, Last updated, Version 1)
XX
DE   S.cerevisiae chromosome XIII cosmid 8156.
XX
KW   BUL1; cytochrome b5; DAG1; initiation factor 1A; inorganic pyrophosphatase;
KW   IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element;
KW   TIF11; TPS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP;
KW   URA10.
XX
OS   Saccharomyces cerevisiae (baker's yeast)
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
XX
RN   [1]
RP   1-29987
RA   Lye G., Churcher C.M.;
RT   ;
RL   Unpublished.
XX
RN   [2]
RP   1-29987
RA   Barrell B., Rajandream M.A.;
RT   ;
RL   Submitted (12-MAY-1995) to the EMBL/GenBank/DDBJ databases.
RL   Saccharomyces cerevisiae chromosome XIII sequencing project, Sanger Centre,
RL   Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrell@sanger.ac.uk
XX
DR   SGD; S0004873; YMR260C.
DR   SGD; S0004874; YMR261C.
DR   SGD; S0004875; YMR262W.
DR   SGD; S0004876; YMR263W.
DR   SGD; S0004877; YMR264W.
DR   SGD; S0004878; YMR265C.
DR   SGD; S0004879; YMR266W.
DR   SGD; S0004880; YMR267W.
DR   SGD; S0004881; YMR268C.
DR   SGD; S0004882; YMR269W.
DR   SGD; S0004883; YMR270C.
DR   SGD; S0004884; YMR271C.
DR   SGD; S0004885; YMR272C.

```

THIS PAGE BLANK (b)(7)(D)

```

DR   SGD; S0004886; YMR273C.
DR   SGD; S0004887; YMR274C.
DR   SGD; S0004888; YMR275C.
DR   SWISS-PROT; P28239; IPY2_YEAST.
DR   SWISS-PROT; P30402; PYRX_YEAST.
DR   SWISS-PROT; P38426; TPS3_YEAST.
DR   SWISS-PROT; P38428; YM8E_YEAST.
DR   SWISS-PROT; P38429; YM8D_YEAST.
DR   SWISS-PROT; P38430; YM8C_YEAST.
DR   SWISS-PROT; P38912; IF1A_YEAST.
DR   SWISS-PROT; P48524; BUL1_YEAST.
DR   SWISS-PROT; P49960; PR24_YEAST.
DR   SWISS-PROT; P50111; ZDS1_YEAST.
DR   SWISS-PROT; P53437; RRN9_YEAST.
DR   SWISS-PROT; Q03496; YM8B_YEAST.
DR   SWISS-PROT; Q03508; YM8F_YEAST.
DR   SWISS-PROT; Q03516; YM8G_YEAST.
DR   SWISS-PROT; Q03525; YM8H_YEAST.
DR   SWISS-PROT; Q03529; SCS7_YEAST.
DR   SWISS-PROT; Q03530; RCE1_YEAST.

```

XX

CC Notes:

XX

CC All CDS over 100 codons have been analysed. CDS that are
CC completely overlapped and those that are overlapped by more than
CC 50%
CC of their length by a larger CDS have been omitted from this
CC analysis.
CC Details of the omitted CDS are available on request.
CC The more significant matches with motifs in the PROSITE database
CC are
CC also included but some of these may be fortuitous.
CC The length in codons and the calculated codon adaptation index
CC (CAI)
CC is given for each CDS.

XX

CC Cosmid 8156 is overlapped at the start by cosmid 9920, embl entry
CC SC9920,
CC accession no. Z48639 and at the end by cosmid 8021.

XX

Key	Location/Qualifiers
source	1..29987 /chromosome="XIII" /db_xref="taxon:4932" /organism="Saccharomyces cerevisiae" /strain="AB972" /clone="cosmid 8156" /map="13R"
misc_feature	1..140 /note="overlap with SC9920 Z48639 S.cerevisiae chromosome XIII cosmid 9920"
CDS	complement(<1..356) /db_xref="SWISS-PROT:Q03496" /note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YM9920.13c" /product="unknown" /protein_id="CAA89242.1" /translation="MTTDVQFSSQEIELFRVKEFLIANNPAKINNENKDAVLTQIEHDF RYLIQYIKDGLPNLNSTRILFPDTFSICLLRSHQIIASKKIDSQEFLSAVKEQLLTEA NANIIFEYVLDFWA"
CDS	complement(851..1312) /db_xref="SGD:S0004873" /db_xref="SWISS-PROT:P38912" /note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, initiation factor 1a" /gene="TIF11"

THIS PAGE BLANK (LSP70)

```

FT      /product="Tif11p"
FT      /protein_id="CAA89243.1"
FT      /translation="MGKKNTKGGKKGRRGKNDSDGPKRELIYKEEGQEYAQITKMLGNG
FT      RVEASCFDGNKMAHIRGKLRKKVWMGQGDIIILVSLRDFQDDQCDVVHKYNLDEARTLK
FT      NQGELPENAKINETDNFGFESDEDVNFEGNADEDEDEGEDEELDIDDI"
FT      CDS      complement(1678..4842)
FT      /db_xref="SGD:S0004874"
FT      /db_xref="SWISS-PROT:P38426"
FT      /note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
FT      SW:TPS3_YEAST P38426; alpha,alpha-trehalose-phosphate
FT      synthase and PIR:S27471; PS00213 Lipocalin signature;
FT      translation in TPS3_YEAST starts at amino acid 33 in this
FT      sequence"
FT      /gene="TPS3"
FT      /product="Tps3p"
FT      /protein_id="CAA89244.1"
FT      /translation="MTIIVASLFLPYTPQFEADVNTSDTAKLVNESSMIKVCNNQELSN
FT      NKQERSSSVTSASSHYIGLPQEAQINGEPLQRANVGSPATGVNYHNEMEMLSSEQFLEE
FT      LTANATHAANS GIPPANPNVSSGSTAQRPSVEEFFSAPSARVCSPSQEASASSISASRS
FT      SAHNDLSSSLMKPNLSFDSHPVRSSSKSAVITPVSKSPDVPDPAVVDVAKVREEF
FT      QQQASLPSMKRVSGSTAGDSSIASSSSNLRYSQQFQDNFIEDTDSIEDIDSDLETDAK
FT      KYNVPKFGGYSNNAKLRLASLMRNSYELFKHLPWTIVDSDKNGSLKNAVNIAVAEKTVK
FT      EPVSVWGTMGIPDELPHVCHKISKKLEQDFSSFPVVTDDITFKGAYKYNQKILWPT
FT      LHYQIPDNPSKAFEDHSWDYYQKVNQKFSRIVSVYKPGDTIWIHDYHMLVPMQVRE
FT      KLPKAKIGFPLHVSFPSEVFRCLANRERILEGIIGANFVGFGQTKEYKRHFLQTCNRL
FT      AADVSNDEVKYHCNIVSVMYAPIGIDYYHLTSQLRNGSVLEWRQLIKERWRNKKLIVCR
FT      DQFDRIRGLQKKMLAYERFLIENPEYIEKVVLIIQICIGKSSDPEYERQIMVVDRINSL
FT      SSNISISQPVVFLHQDLDAQYLALNCEADVFLVDALREGMNLTCHEFIVSSFENAPL
FT      LLSEFTGSSSVLKEGAILINPWDINHVAQSIKRSLEMSPEEKRRRWKKLFKSVEIHDSD
FT      NWITKCFEYINNAWESNQETSTVFNLAPKFCADYKASKKHLFIKISEPPTSRMLSLL
FT      SELSSNNIVYVLSSFTKNTFESLYNGVLNIGLIAENGAYVRVNGSWYNIVEELDWMKEV
FT      AKIFDEKVERLPGSYYKIADSMIRFHTENADDQDRVPTVIGEATHINTLFDDRDHAY
FT      VHKDIVFVQQTGLALAAAEFLMKFYNSGVSPDTSRISLRTSSSSSVGNNKKHFQONQV
FT      DFVCSVGSTSPIIEPLFKLVKQVEKNNLKFYGTILYGSRRSTYAKEHINGVNELFTIL
FT      HDLTAA"
FT      misc_feature      complement(2260..2301)
FT      /note="PS00213 Lipocalin signature"
FT      CDS      5199..6140
FT      /db_xref="SGD:S0004875"
FT      /db_xref="SWISS-PROT:P38430"
FT      /note="YM8156.04, unknown, len: 313, CAI: 0.11, previously
FT      noted ORF, SW:YTP3_YEAST P38430"
FT      /product="unknown"
FT      /protein_id="CAA89245.1"
FT      /translation="MNKLVDACHVITDPDNTFCGDDGGSQGTLCVMSSNPYDWNNLK
FT      KLAGRSTSKNDICVGFVHPWYSHLFYVGSRDKVSHYQDVLEYKNEEQFDSLQVLPPE
FT      PLDLEEYIKREFNDTLVSVIGEIGLDKFLRLPANGFYMONEKARLTTVKVKLSHQETVF
FT      RRFWRRLARHTSKPISIHDKCHGKLNLDICNEELLTYHSVKICLHSYTGSKETLLGQWLK
FT      KFPPDRIFVSLSKWINFKDPEEGDALVRSLPSTCILTETDYPIDNPDPYQKALTEQLQ
FT      YLNAQIARAWDETLDASQAALRVYENFQKFIK"
FT      CDS      6392..6997
FT      /db_xref="SGD:S0004876"
FT      /db_xref="SWISS-PROT:P38429"
FT      /note="YM8156.05, unknown, len: 301, CAI: 0.12,
FT      SW:YTP2_YEAST P38429 previously noted ORF"
FT      /product="unknown"
FT      /protein_id="CAA89246.1"
FT      /translation="MARPVNTNAETESRGRPTQGGGYASNNNGSCNNNNNGSNNNNNNNN
FT      NNNNNNSNNNNNGPTSSGRNTGKQRLTAAQQQYIKNLIETHITDNHPDLRPKSHPMDF
FT      EEYTD AFLRRYKDHFLQDVPDNLTLQGYLLGSKLGAKTYSYKRNTQGQHDKRIHKRDLA
FT      NVVRRHFDEHSIKETDCIPQFIYKVNQKKKFKMEFRG"
FT      CDS      7278..7889
FT      /db_xref="SGD:S0004877"
FT      /db_xref="SWISS-PROT:P38428"
FT      /note="YM8156.06, unknown, len: 301, CAI: 0.13,
FT      SW:YTP1_YEAST P38428, extends previously noted ORF"
FT      /product="unknown"

```

THIS PAGE BLANK (3PT0)


```

FT      /protein_id="CAA89247.1"
FT      /translation="MEDSRLLITLILVFGVIFLKKFFQSNQHPSAQRLSATGVNAHGRP
FT      QGSTQNALRRRTGRVNGGHPVTTQMVETVQNLAPNLHPEQIRYSLENTGSVEETVERYLR
FT      GDEFSFPPGFEPSPRAPMGANAAVDNNAAGGGEFNDPRKKNMICAENLLDKFHVLDLNEDM
FT      SNLSFKDLIDIEERKRLLVWQARKNLETKLQSDKDLQSLLT"
FT      CDS      complement(8013..9398)
FT      /db_xref="SGD:S0004878"
FT      /db_xref="SWISS-PROT:Q03508"
FT      /note="YM8156.07c, unknown, len: 461, CAI: 0.11"
FT      /product="unknown"
FT      /protein_id="CAA89248.1"
FT      /translation="MEEFEEFRKKGEMSSRCGNHRVLRKWNSCACELAVPFEVPEHAIT
FT      KLHIYDFDNTLFATPGPTEQLYTRELLNLLTSSTLPNGGWNEPGFLQAAIEISKTKPR
FT      RYSWNADIVKLAEEESYAKDTISIVLTGREESKFHKLIEHALQTARSHWKCSENEFRFN
FT      AVCLKKRAISEYTSKYKKELMRDFLEYYPSSLRELSIYDDRIHQIDAFKSFFHSLDLPLR
FT      KWSAIPVRPFTKSLPREQELEMVMDMVRKNNSQALSTSQKFDLRRTPRQIGYILCTASH
FT      RLLSIEVIKYLKRRKGRRTFRPKLYEHPLYIPCAEPGKDIPALEIAKVWSNNDTRTFDS
FT      EKKVQHISQIFYLEQPGKCIVHFQVTDLAVIASAHNRRKPLEVYFKATPEPNRYTFTL
FT      FPYIVTGHFYKRDRIEDLEVVTTERLINCKEDIHWVPLDNTIPIKAFFGRFAKLAAIPC
FT      SNA"
FT      CDS      9991..12852
FT      /db_xref="SGD:S0004879"
FT      /db_xref="SWISS-PROT:Q03516"
FT      /note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
FT      integral membrane glycoprotein"
FT      /product="unknown"
FT      /protein_id="CAA89249.1"
FT      /translation="MNSTNSTNSTTTATSTNTSTQQVVTSLVSNGTIFGVFVIAFLILR
FT      IKLKRIYEPKSSFNLINEEKKPEPLPQGVWQWLKPLLKKSDNFVIQQAGLDGYFFLRYL
FT      FIIAIYCAVMSYIFPILLSINASNGNHESGLNQLAYQNVKHRGRYFAHVFCGWIFFWG
FT      FLYIIYRELYFYTSMKQAVLASPRYAKKLSSRTVLFQTVPKQYLSEEEFSKLFQDGVKRV
FT      WIARGSGSIEAMVKARDNMAIQLEGAETKYLKAALKKKIKLNLKKSQPLSVSDNIAEYVP
FT      DKKRPHHKINKVAKFFFGKKVDTISYIKEELPKLNQKVKALQEDHENSPPNSVFVEFE
FT      SQYQAQVAAQITTYHAPLFMTPVYIGIEPSDVVWFNLRMFWWERLGREVSAVSAIVALV
FT      ILWAFPVAVGFMISNITSLTNEVKWLKFIYKLPKQLLGLLTSAPTVALAVLMSFLPKF
FT      IRGMAITQGAPSKQVVEYFTQAYFAFQVIQVFLVTTLSAATSTVTEIVKEPTKAMD
FT      LASNLPKASNPFMSYVILQGLSISGALLQIVPLILFYVLGAFDGTVRKKWNRFGLS
FT      SMQWGTAFFPVYTNLAVITFSYSIISPLILLFAAVAFFLLYIAYLYNLTYVYQESPDARG
FT      IYYPRALFQITVIGIYIGQICLLGLFAVGKGGPIVLQVIGICVTVLIHLHLSAAPDHL
FT      KVIPVDTMKPLDGVSDTPSFKNYKGIESTKVKKNTFGANIDMDGKELPEFPKIKYHK
FT      RSESVTEQQVENSIFSENTFEYQFNPAEANADGHAINAENLIEDVPLLADGDTMKIPP
FT      APWWKRFLKPHIYYSYKAVKSRLPEIYGLVDPDERVNDFDISHAYDPAVSAQCPDWI
FT      PRDPFGFSKLLISDVSGVVMNDENATIDENLKFTLRDVPPPYNVVKDEANGEANGEDF
FT      TASKENNPADPKYKEEESRSV"
FT      CDS      13245..14177
FT      /db_xref="SGD:S0004880"
FT      /db_xref="SWISS-PROT:P28239"
FT      /note="YM8156.09, IPP2 gene (PPA2), len: 310, CAI: 0.14,
FT      SW:IPY2_YEAST, P28239, inorganic pyrophosphatase,
FT      mitochondrial; contains PS00387 Inorganic pyrophosphatase
FT      signature"
FT      /gene="IPP2"
FT      /product="Ipp2p"
FT      /protein_id="CAA89250.1"
FT      /translation="MNLRLMNAITSKARSIERLQTLNLSIRNHRQFSTIQQGSKYTL
FT      GFKKYLTLNGEVGSFFHDVPLDLNEHEKTVNMIVEVPRWTTGKFEISKELRFNPVQD
FT      TKNGKLRFVNIPYHGYIHNYGAIPQTWEDPTIEHKLKCDVALKGDNDPLDCCEIGS
FT      DVLEMGSIKKVKVLGSLALIDDGELDWKVIVIDVNDPLSSKIDDLKIEEYFPGILD
FT      REWFRKYKVPAGKPLNSFAFHEQYQNSNKTIQTIKKCHNSWKNLISGSLQEKYDNLNPT
FT      ERAGNGVTLEDVSVKPPSQIPPEVQKWWYV"
FT      misc_feature 13698..13718
FT      /note="PS00387 Inorganic pyrophosphatase signature"
FT      CDS      complement(14361..15695)
FT      /db_xref="SGD:S0004881"
FT      /db_xref="SWISS-PROT:P49960"
FT      /note="YP8156.10c, U6 snRNP gene, len: 444, CAI: 0.12,
FT      amino acids 200-288 identical to PIR:A39562 A39562 U6 snRNP

```

THIS PAGE BLANK (USPTO)

FT protein PRP24"
 FT /gene="U6sNRP"
 FT /product="U6sNRpP"
 FT /protein_id="CAA89251.1"
 FT /translation="MEYGHHPDPSKRPLDEGSPAAAGLTSSKANEALTRNRELTTVLV
 FT KNLPKSYNQNKVYKYFKHCGPIIHVDVADSLKKNFRFARIEFARYDGALAAITKTHKV
 FT GQNEIIIVSHLTECTLWMTNFPSPYTQRNIRDLLQDINVVALSIRLPSLRFNTSRRFAYI
 FT DVTSKEDARYCVEKLNGLKIEGYTLVTKVSNPLEKSKRTDSATLEGREIMIRNLSTELL
 FT DENLLRESFEGFGSIEKINIPAGQKEHSFNNCCAFMVFENKDSAERALQMNRSLLGNRE
 FT ISVSLADKKPFLERNEVKRLLASRNSKELETLICLFPLSDKVSPSLICQFLQEEIHINE
 FT KDIRKILLVSDFNCAIIIFRDSKFAAKMLMILNGSQFQGVIRSGTINDMKRYNNQGN
 FT HSMKHVKPSCINMMEKGNLQVKKKIPDKQEOMSNDDFRKMFLGE"
 FT CDS 16137..16565
 FT /db_xref="SGD:S0004882"
 FT /db_xref="SWISS-PROT:Q03525"
 FT /note="YM8156.11, unknown, len: 142, CAI: 0.13, Lys-, Ser-
 FT rich hypothetical protein"
 FT /product="unknown"
 FT /protein_id="CAA89252.1"
 FT /translation="MAVLNLLQNEAVATAVSKSSSPLYRWFVKGEGLKGTITNLGKKEE
 FT ASFVVSASSSSKGKKRRRDEDDNKVKRKKLKKDKKTSNDSESCKKKKKKSKESKKGK
 FT KSKHSSDEGDKSKHKSSKSSKHKHKEESSARRDRKEHI"
 FT CDS complement(16800..17897)
 FT /db_xref="SGD:S0004883"
 FT /db_xref="SWISS-PROT:P53437"
 FT /note="YM8156.12c, unknown, len: 365, CAI: 0.14"
 FT /product="unknown"
 FT /protein_id="CAA89253.1"
 FT /translation="MSDLDEESQIETQIDAPIEDIIRGSELTTTADKETLKSANELLD
 FT SLEHSHRVDLSLHLYSAYLLKRLLYKANEEKHFYEVNQFVKQIKDNWTSWPNPNTIID
 FT PSVDKLYEDIPEGIANVSVQGEISNRALMHASDMMRVELDAQWQKFLSKSALDHDVTL
 FT DVDELNIPNEISRNILVKLDSLFEGLHDKIAKENEFQVQDKHSNNIRANQIDDEPMQA
 FT NRRRIKYTYHDLVSRGCEMNEDMTDIYMKSLLEYNDIPEKYKKRKFRLPKQILKKYHQP
 FT KTSSYLKELLKSTREDFIPVEKLLKDKRLTSKDKSKLQRLNREETEDALNKRTPFFQVKG
 FT YLEDENEISDYELDDCLIELPNGNI"
 FT CDS complement(18339..19022)
 FT /db_xref="SGD:S0004884"
 FT /db_xref="SWISS-PROT:P30402"
 FT /note="YM8156.13c, URA10, len: 227, CAI: 0.14,
 FT SW:PYRX_YEAST P30402, contains orotate
 FT phosphoribosyltransferase 2; PS00103 Purine/pyrimidine
 FT phosphoribosyl transferases signature"
 FT /gene="URA10"
 FT /product="Ura10p"
 FT /protein_id="CAA89254.1"
 FT /translation="MSASTTSLEEYQKTFLELGLGLECKALRFGSFKLNSGRQSPYFFNLS
 FT LFNSGKLLANLATAYATAIIQSELKFDVIFGPAYKGIPLAAIVCVKLAIEIGGTFQGIQ
 FT YAFNRKKVKDHGEGGIIVGASLEDKRVLIIDVMTAGTAINEAFEIISIAQGRVVGCV
 FT ALDRQEVIHESDPERTSATQSVSKRYNVPVLSIVSLTQVVQFMGNRLSPEQKSAIENYR
 FT KAYGI"
 FT misc_feature complement(18594..18632)
 FT /note="PS00103 Purine/pyrimidine phosphoribosyl
 FT transferases signature"
 FT tRNA complement(19720..19791)
 FT /note="tRNA-Gln gene, anticodon ctg, len: 72"
 FT /gene="tRNA-Gln"
 FT LTR complement(20011..20380)
 FT /note="solo tau element"
 FT CDS complement(21097..22251)
 FT /db_xref="SGD:S0004885"
 FT /db_xref="SWISS-PROT:Q03529"
 FT /note="YM8156.14c, unknown, len: 384, CAI: 0.25, similar to
 FT cytochrome b5/nitrate reductase family; PS00191 Cytochrome
 FT b5 family, heme-binding domain signature"
 FT /product="unknown"
 FT /protein_id="CAA89255.1"
 FT /translation="MSTNTSKTLELFSKKTQVEHNTANDCWVTYQNRKIYDVTRFLSEH

THIS PAGE BLANK (0.3870)

```

FT      PGGDESILDYAGKDITEIMKDSVDVHEHSDSAYEILEDEYLLIGYLATDEEAARLLTNKNH
FT      KVEVQLSADGTEFDSTTFVKELPAEEKLSIATDYSNDYKHKHFLDLNRPLLMQILRSDF
FT      KKDFYVDQIHRPRHYGKGSAPLFGNFLEPLTKTAWVVPVAVLPPVVVYHMGVALKNMNQ
FT      LFACFLFCVGVFVWTLIEYGLHRFLFHFDLWLPESNIAFATHFLLHGCHHYLPMDKYRL
FT      VMPPTLFLVILCAPFYKLVFALLPLYWAYAGFAGGLFGYVCYDECHFFLHHSKLPFPMRK
FT      LKKYHLEHHYKNYQLGFGVTSWFWDDEVFGTYLGPDAPLSKMKYE"
FT      misc_feature complement(22108..22131)
FT      /note="PS00191 Cytochrome b5 family, heme-binding domain
FT      signature"
FT      CDS complement(22707..25454)
FT      /db_xref="SGD:S0004886"
FT      /db_xref="SWISS-PROT:P50111"
FT      /note="YM8156.15c, unknown, len: 915, CAI: 0.14"
FT      /product="unknown"
FT      /protein_id="CAA89256.1"
FT      /translation="MSNRDNESMLRTTSSDKAIASQRDKRKSEVLIAAQSLDNEIRSVK
FT      NLKRLSIGSMDLLIDPELDIKFGGESSGRRSWGTTSSSASMPSDTTTNNTRYSDPTP
FT      LENLHGRGNSGIESSNKTQGNYLGIKKGVHSPSRKLNANVLKKNLLWVPANQHPNVKP
FT      DNFLVLVQDTLQNIQLSDNGEDNDGNSNENNDIEDNGEDKESQSYENKENNTINLNRGL
FT      SRHGNASLIRRPSTLRRSYTEFDDNEDDDNKGDSASETVNKVEERISKIKERPVSRLDI
FT      TEELTKISNSAGLTDNDAITLARTLSMAGSYSDKKDQPQPEGHYDEGDIGFSTSQANTL
FT      DDGEFASNMPINNMTWPERSSLRRSRFRNTYRIRSQEKEVEQSVDEMKNDDERLKL
FT      TKNTIKVEIDPHKSPFRQQDESDENMSSPGSIGDFQDIYNHYRQSSGEWEQEMGIEKEA
FT      EEPVVKVRNDTEVQDLELREGTTDMVKPSATDDNKETKRHRRRNGWTWLNKMSREDDN
FT      EENQDDENEENEVDQSQRMELDNSKKHYISLFNGGEKTEVSNKEEMNSSTSTATSQTRQ
FT      KIEKTFANLFRRKPHHKHDASSPSSSPSSSPSIPNNDVHVRVRKSKKLGKNSGREPV
FT      EPIVLRNRPRPHRHHHSRHGSQKISVKTLKDSQPQQQIPLQPQLEGAIEIEKKEESDSE
FT      SLPQLQPAVSVSSTKSNSRDREEEAAKKKNKRSNTTEISNQQHSKHVQKENTDEQKAQ
FT      LQAPAQEQVQTSVPVQASAPVQNSAPVQTSAPVEASAQTQAPAAPPLKHTSILPPRKLT
FT      FADVKKPKDKNPSVPQFTDSAFGFPLPLLTVSTVIMFDHRLPINVERAIYRLSHLKLNSN
FT      KRGLEQVLLSNFMYAYLNLVNHTLYMEQVAHDKEQQQQQQQQP"
FT      CDS complement(25838..26785)
FT      /db_xref="SGD:S0004887"
FT      /db_xref="SWISS-PROT:Q03530"
FT      /note="YM8156.16c, unknown, len: 315, CAI: 0.10"
FT      /product="unknown"
FT      /protein_id="CAA89257.1"
FT      /translation="MLQFSTFLVLLYISISYVLPLYATSQPEGSKRDNPRTIKSRMQKL
FT      TIMLISNLFVLPFLQSQLSSTTSHISFKDAFLGLGIIPGYAALPNPWQFSQFVKDLTK
FT      CVAMLLTLYCGPVLDVFLYHLLNPKSSILEDFYHEFLNIWSFRNFIFAPITEEIFYTSM
FT      LLTTYLNLIPHSQLSYQQFLWQPSLFFGLAHAAHAYEQLQEGSMTTVSILLTTCFQILY
FT      TTLFGGLTKFVFVRTGGNLWCCIIILHALCNIMFGPGPSRLNLHFTVVDKKAGRISKLV
FT      IWNKCYFALLVLGLISLKDTLQTLVGTPGYRITL"
FT      CDS complement(27125..>29987)
FT      /codon_start=2
FT      /db_xref="SGD:S0004888"
FT      /db_xref="SWISS-PROT:P48524"
FT      /note="YM8156.17c, BUL1 gene, (DAG1), len: >953, CAI: 0.14,
FT      EM_NEW:SCBUL1 D50083"
FT      /gene="BUL1"
FT      /product="Bul1p"
FT      /protein_id="CAA89258.1"
FT      /translation="DFTANSSTTMNVNANTRGRGRQKQEGGKSSRSPLHSPKSWIRS
FT      ASATGILGLRRPELAHSHSHAPSTGTGAGNRSPLRRSTANATPVETGRSLTDGDIINV
FT      VDVLPSEFMYNTLHRHIPQGNVDPDRHDFPPSYQEANNSTATGAAGSSADLSHQSLSTD
FT      ALGATRSSSTSNLENLIPLRTEHHSIAAHQSTAVDEDSLDIPPILLDDLNDDNIFIDKL
FT      YTLPKMSTPIBITIKTKHAPIPHVKPEEESILKEYTSGDLIHGFIITENKSQANLKEF
FT      MFYVTLESYISIIDKVKSKRTIKRFLRMVDLSASWSYSKIALGSGVDIFPADVDYDGSV
FT      FGLNNSRVLEPGVKYKFFIFKLPQLQLLDVTCKQEHFSSHCLLPSPSGIDKYRNNCKYSG
FT      IKVNRVLGCGHLGTGSPILTNDMSDDNLSINYTIDARIVGKDQKASKLYIMKEREYNL
FT      RVIPFGFDANVVGERTTMSQLNDITKLVOERLDALRKIFORLEKKEPITNRDIHGADLS
FT      GTIDDSIESDSQEIQLQRKLDQLHIKNRNNYLVNYNDLKLGHDLNDRSGNSGHNTDTSR
FT      AWGPFVESELKYKLNKNSSSSFLNFSHFLNSSSSSSMSSSSNAGKNNHDLTGKNKERTGL
FT      ILVKAKIPKQGLPYWAPSLLRKTNVFESKSKHDQENWVRLSELIPEDVKKPLEKLDLQL
FT      TCIESDNLSPHDPPEIQSITTELCITAKSDNSIPIKLNSELLMNKEKLTISKALYDDF
FT      HSKICEYETKFNKNFLELNELYNMNRGDRRPKELKFTDFITSQLFNDIESICNLKVSFH
FT      NLSNIFKKQVSTLQHSKHALSEDSISHTGNGSSSSPSSASLTPVTSSSKSSLFLPSGS

```

THIS PAGE BLANK (1-370)

FT SSTS LKFTDQIVHKWVRIAPLQYKRDINVNLEFNKDIKETLIPSFESCLCCRFYCVRVM
FT IKFENHLGVAKIDIPISVRQVTK"
FT misc_feature 29849..29987
FT /note="overlap with cosmid 8021"
XX
SQ Sequence 29987 BP; 8850 A; 5706 C; 5896 G; 9535 T; 0 other;

Sequence omitted for brevity

Updates - The EMBL Nucleotide Sequence Database
EMBL Outstation Hinxton, The European Bioinformatics Institute,
Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, U.K.
Tel: +44 1223 494499. Fax: +44 1223 494472

Internet electronic addresses:
datasubs@ebi.ac.uk (data submissions), update@ebi.ac.uk (updates)
datalib@ebi.ac.uk (general enquiries)
<http://www.ebi.ac.uk/> (world wide web submissions, updates, services,
info)

THIS PAGE BLANK (JPTO)

RP CHARACTERIZATION.
 RX MEDLINE-20285442; PubMed-10825201;
 RA Trublood C.E., Boyartchuk V.L., Pictologlou E.A., Rozema D.,
 RA Poulter C.D., Rine J.,
 RT "The Caax proteases, Atp1p and Rce1p, have overlapping but distinct
 RT substrate specificities."
 RT Mol. Cell. Biol. 20:4381-4392(2000).
 RN (8)
 RP SUBCELLULAR LOCATION.
 RX MEDLINE-98409630; PubMed-9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.,
 RT "Endoplasmic reticulum membrane localization of Rce1p and Ste24p.
 RT Yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal a-factor cleavage."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR BINDING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249617; CA89647.1;
 DR EMBL: 077137; AAB38271.1;
 DR MEROPS: M48.001;
 DR SGD: S0003878; STE24.
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum; Pheromone response.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT METAL 297 297 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 298 298 BY SIMILARITY.
 FT METAL 301 301 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 390 390 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 394 394 PHOTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 453 AA; 52324 MW; 331CC9AE2D7C99DA CRC64;

Query Match 99.7%; Score 2336; DB 1; Length 453;
 Best Local Similarity 99.8%; Pred. No. 3.7e-160;
 Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MEDLKTILDPNIPKLIISGFSIAOFSESYLTROYKLSKLPVLEDEIDDETH 60
 DB 1 MEDLKTILDPNIPKLIISGFSIAOFSESYLTROYKLSKLPVLEDEIDDETH 60
 OY 61 KSNINIRAKAKSIFGDVYNLAOKLVFIYDLPFKIMHNAVSLNAVLPVRHMSTVAQ 120
 DB 61 KSNINIRAKAKSIFGDVYNLAOKLVFIYDLPFKIMHNAVSLNAVLPVRHMSTVAQ 120
 OY 121 SICFTGLSSITLVDPSTSYSHFVLEKFGFNKLTQVLMITDMKSLTLAAGCPTL 180
 DB 121 SICFTGLSSITLVDPSTSYSHFVLEKFGFNKLTQVLMITDMKSLTLAAGCPTL 180

OY 181 YLEKIPDKPEPTFLMYINVFVQIILAMTIIPIVIMPMFKPTPLEDEKLSIESLA 240
 DB 181 YLEKIPDKPEPTFLMYINVFVQIILAMTIIPIVIMPMFKPTPLEDEKLSIESLA 240
 OY 241 DRVGFPLKIFVIDSKRSSHNAFYTGLPFSKRIVLEDTLVNSNSTDEITAVLAHEIG 300
 DB 241 DRVGFPLKIFVIDSKRSSHNAFYTGLPFSKRIVLEDTLVNSNSTDEITAVLAHEIG 300
 OY 301 HMOKHIVNMVIFSQLHTFLIFSLTSTIYRNTSEYNTGFELEKSTGSFVDVITKEPPI 360
 DB 301 HMOKHIVNMVIFSQLHTFLIFSLTSTIYRNTSEYNTGFELEKSTGSFVDVITKEPPI 360
 OY 361 IIGFNLFDLTPLECANQFVMSLSTRHEVADAVAKKLGKONLCRALDLOIKNST 420
 DB 361 IIGFNLFDLTPLECANQFVMSLSTRHEVADAVAKKLGKONLCRALDLOIKNST 420
 OY 421 MNVDPLSSYHSHPTLERSALDYSEKKKN 453
 DB 421 MNVDPLSSYHSHPTLERSALDYSEKKKN 453
 DB 421 MNVDPLSSYHSHPTLERSALDYSEKKKN 453
 RESULT 2
 ST24_SCHPO STANDARD: PRT: 474 AA.
 ID ST24_SCHPO
 AC Q10071.
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-
 DE SPECIFIC ENDOPEPTIDASE 1) (PPSEP 1).
 GN SPAC3H1.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Gentiles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED PROTEINS (BY SIMILARITY).
 CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 268144; CA92258.1;
 DR MEROPS: M48.001;
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 230 250 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT METAL 332 332 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 333 333 BY SIMILARITY.
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342
Sequence: 1 MEDLKITLIDHPNIPMKLIIS.....HPTLAERSTALDYVSEKKKN 453

Scoring table:

BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2336	99.7	453	1 ST24_YEAST	P47154 saccharomyc
2	903	38.6	474	1 ST24_SCHPO	010071 schizosacch
3	698.5	29.8	475	1 FAC1_HUMAN	075844 homo sapien
4	328	14.0	426	1 YHPN_BACSU	P40769 bacillus su
5	161	6.9	286	1 HTPX_PASMU	P57846 pasteurella
6	150.5	6.4	310	1 HTPX_HELPJ	092854 helicobacte
7	150	6.4	335	1 HTPX_ARCFU	030004 archaeoglob
8	147.5	6.3	283	1 HTPX_HAEIN	P44840 haemophilus
9	142.5	6.1	310	1 HTPX_HELPY	025882 helicobacte
10	139	5.9	366	1 HTPX_AERPE	094667 aeropyrum p
11	138.5	5.9	258	1 HTPX_METTH	026869 methanobact
12	130.5	5.6	292	1 HTPX_PYRHO	058997 pyrococcus
13	130.5	5.6	317	1 HTPX_THERC	094172 thermoplasma
14	127.5	5.4	2136	1 YCF2_MARPO	P09979 marchantia
15	120	5.1	292	1 HTPX_BUCAL	P57406 buchnera ap
16	119.5	5.1	2294	1 YCF2_ARATH	P56786 arabidopsis
17	119	5.0	297	1 HTPX_STRGC	030795 streptococc
18	116	5.0	293	1 HTPX_ECOLI	P23894 escherichia
19	113	4.8	284	1 HTPX_METJA	059076 methanococc
20	113	4.8	298	1 HTPX_BACSU	031657 bacillus su
21	109	4.7	506	1 YCF2_EUGRA	P31921 euglena gra
22	109	4.7	595	1 YCF6_METJA	058682 methanococc
23	107.5	4.6	1032	1 MT18_YEAST	P40469 saccharomyc
24	106.5	4.5	646	1 NTP1_HAEPV	037319 heliothis a
25	106.5	4.5	1029	1 YFC5_YEAST	P43571 saccharomyc
26	105.5	4.5	873	1 COX1_ACACA	037370 acanthamoeb
27	105	4.5	1427	1 SRB8_YEAST	P25648 saccharomyc
28	104.5	4.5	585	1 YHD9_YEAST	P38732 saccharomyc
29	104.5	4.5	1100	1 LOS1_YEAST	P33418 saccharomyc
30	104	4.4	302	1 HTPX_AQUAE	067798 aquifex aeo
31	104	4.4	530	1 CP14_CHICK	P79760 gallus gall
32	103.5	4.4	185	1 GU01_RAT	P35994 rattus norv
33	103.5	4.4	450	1 VATA_SCHPO	014265 schizosacch

34	103	4.4	645	1 VP74_NPYAC	P15963 autographa
35	103	4.4	1101	1 KOE5_YEAST	Q08217 saccharomyc
36	102	4.4	577	1 YHT1_SCHPO	Q09684 schizosacch
37	102	4.4	2493	1 YBA4_YEAST	P35194 saccharomyc
38	100.5	4.3	260	1 YMI6_BRECAM	Q21266 recinnomona
39	100.5	4.3	1040	1 RIK1_SCHPO	Q10426 schizosacch
40	100	4.3	1325	1 MRP4_HUMAN	Q15439 homo sapien
41	99.5	4.2	309	1 O7AA_HUMAN	Q76100 homo sapien
42	99	4.2	485	1 YC11_KLEPN	Q48457 klebsiella
43	99	4.2	1024	1 RPOB_PLAFA	P21421 plasmodium
44	98.5	4.2	313	1 ND2M_RHISA	Q09817 rhipicephal
45	98.5	4.2	417	1 O65A_DROME	P82982 drosophila

ALIGNMENTS

RESULT	ID	ST24_YEAST	STANDARD:	PRT:	453 AA.
AC	P47154:	ST24_YEAST			
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	CAAX PRENYL PROTEIN 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC				
DE	ENDOPROTEINASE 1) (PSEF 1) (A-FACTOR CONVERTING ENZYME).				
GN	ST24 OR AFI1 OR YJRL17M OR J2032.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N. A.				
RA	Rose M., Koelter P., Entian K. D.;				
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N. A., AND FUNCTION.				
RX	MEDLINE=97167681; PubMed=9015299;				
RA	Fujimura-Kamada K., Nouvet F. J., Michaelis S.;				
RT	"A novel membrane-associated metalloprotease, Ste24p, is required for				
RT	the first step of NH2-terminal processing of the yeast a-factor				
RT	precursor.";				
RL	J. Cell Biol. 136:271-285(1997).				
RN	[3]				
RP	FUNCTION.				
RX	MEDLINE=97218305; PubMed=9065405;				
RA	Boyarichuk V. L., Ashby M. N., Rine J.;				
RT	"Modulation of Ras and a-factor function by carboxyl-terminal				
RT	proteolysis.";				
RL	Science 275:1796-1800(1997).				
RN	[4]				
RP	FUNCTION.				
RX	MEDLINE=98365461; PubMed=9700155;				
RA	Tam A., Nouvet F. J., Fujimura-Kamada K., Slunt H., Sisodia S. S.;				
RT	"Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal				
RT	proteolysis and COOH-terminal CAAX processing.";				
RL	J. Cell Biol. 142:635-649(1998).				
RN	[5]				
RP	FUNCTION.				
RX	MEDLINE=98393572; PubMed=9725832;				
RA	Boyarichuk V. L., Rine J.;				
RT	"Roles of prenyl protein proteases in maturation of Saccharomyces				
RT	cerevisiae a-factor.";				
RL	Genetics 150:95-101(1998).				
RN	[6]				
RP	FUNCTION.				
RX	MEDLINE=20158928; PubMed=10692417;				
RA	Schmidt W. K., Tam A., Michaelis S.;				
RT	"Reconstitution of the Ste24p-dependent N-terminal proteolytic step in				
RT	yeast a-factor biogenesis.";				
RL	J. Biol. Chem. 275:6227-6233(2000).				
RN	[7]				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(without alignments) 10520.305 Million cell updates/sec

Title: US-09-165-460A-3
2948

Perfect score: 1 tgaactgttgatgacaagaag.....ggggagagataagaatcaca 2948

Sequence: 1 tgaactgttgatgacaagaag.....ggggagagataagaatcaca 2948

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb.ba:*
- 2: gb.htg:*
- 3: gb.in:*
- 4: gb.om:*
- 5: gb.ov:*
- 6: gb.pat:*
- 7: gb.ph:*
- 8: gb.pl:*
- 9: gb.pr:*
- 10: gb.ro:*
- 11: gb.sts:*
- 12: gb.sy:*
- 13: gb.un:*
- 14: gb.vl:*
- 15: em.ba:*
- 16: em.fun:*
- 17: em.hum:*
- 18: em.in:*
- 19: em.om:*
- 20: em.or:*
- 21: em.ov:*
- 22: em.pat:*
- 23: em.ph:*
- 24: em.pl:*
- 25: em.ro:*
- 26: em.sts:*
- 27: em.sy:*
- 28: em.un:*
- 29: em.vl:*
- 30: em.htgo.hum:*
- 31: em.htgo.inv:*
- 32: em.htgo.rod:*
- 33: em.htg.hum:*
- 34: em.htg.inv:*
- 35: em.htg.rod:*
- 36: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	2948	100.0	29987	8	SCB156	249260 S. cerevisiae
2	1560.4	52.9	4156	8	D50276	D50276 Saccharomyces
3	1361	46.2	4130	8	YSCNCR1G	142891 Saccharomyces
4	1163	39.5	4225	8	SCU56882	U56882 Saccharomyces
5	1069.4	36.3	3250	8	SCU32580	U32580 Saccharomyces
6	931	31.6	3525	8	SCU63849	U63849 Saccharomyces
7	786.8	26.7	3240	8	YSCBUL1	D50890 Saccharomyces
8	703.8	23.9	3020	8	SCROSDIGEN	X88901 S. cerevisiae
9	225.2	7.6	270	8	YSCGNC58X	M87379 Yeast Eco R
10	143.8	4.9	27559	8	SCB339	249210 S. cerevisiae
C 11	71.8	2.3	2023	8	SCJNR066C	U13683 S. cerevisiae
C 12	68	2.4	4775	8	SCU32938	U32938 Saccharomyces
C 13	65.6	2.3	780	11	CNS06K8B	AL040561 T7 end of
C 14	65.6	2.2	256172	2	AC005139	AC005139 Plasmid
C 15	64.4	2.2	310779	2	AC005140	AC005140 Plasmid
C 16	64.4	2.2	904	11	CNS06K8B	AL039867 T3 end of
C 17	64.4	2.2	86827	3	PFMAL3P5	AL034556 Plasmid
C 18	63.4	2.2	1920	3	PFMAL3P7	X62393 P. falciparum
C 19	63.4	2.2	110000	2	PFMAL13P2_0	AL049185 Plasmid
C 20	62	2.1	253305	3	PFMAL3P7	AL034559 Plasmid
C 21	61.6	2.1	1141	6	AX083744	AL034575 Plasmid
C 22	60.8	2.1	224112	3	PFMAL4P2	AL034575 Plasmid
C 23	60.6	2.1	202872	2	AC016160	AC016160 Homo sapi
C 24	60	2.0	5632	3	AF310892	AF310892 Dictyostea
C 25	59.4	2.0	3214	3	AF019980	AF019980 Dictyostea
C 26	59.4	2.0	318221	2	PFMAL13P3	AL049184 Plasmid
C 27	58.8	2.0	321003	2	PFMAL4P3	AL035476 Plasmid
C 28	58.4	2.0	78439	17	HS085157	U85157 Homo sapien
C 29	58.4	2.0	169546	2	AC004157	AC004157 Plasmid
C 30	58.4	2.0	245802	2	AC006279	AC006279 Plasmid
C 31	58.4	2.0	251124	2	HUAE000660	AE000660 Homo sapi
C 32	58.2	2.0	153098	3	PFMAL3P2	AL034558 Plasmid
C 33	58	2.0	169794	2	AC004688	AC004688 Plasmid
C 34	58	2.0	178137	9	AC011302	AC011302 Homo sapi
C 35	57.8	2.0	13684	3	AE001403	AE001403 Plasmid
C 36	57.8	2.0	153477	2	AC006278	AC006278 Plasmid
C 37	57.6	2.0	53932	2	AC023371	AC023371 Homo sapi
C 38	57	1.9	12029	3	AE001431	AE001431 Plasmid
C 39	57	1.9	153443	2	AC006280	AC006280 Plasmid
C 40	57	1.9	178273	2	AC005308	AC005308 Plasmid
C 41	56.8	1.9	166149	2	AC004709	AC004709 Plasmid
C 42	56.8	1.9	67970	3	PFMAL1P3	AL031746 Plasmid
C 43	56.8	1.9	149627	9	AC087428	AC087428 Homo sapi
C 44	56.6	1.9	92633	2	PFMAL4P1_3	Continuation (4 of
C 45	56.6	1.9	153477	2	AC006278	AC006278 Plasmid

ALIGNMENTS

RESULT 1

LOCUS SCB156 29987 bp DNA PLN 11-AUG-1997

DEFINITION S. cerevisiae chromosome XIII cosmid 8156.

ACCESSION 249260 271257

VERSION 249260.1 GI:809081

KEYWORDS BUL1, cytochrome b5; DAG1, initiation factor 1A; inorganic pyrophosphatase; IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF1; TRS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP; URA10.

SOURCE Baker's yeast.

ORGANISM Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 29987)

AUTHORS Lye, G. and Churcher, C. M.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 29987)

AUTHORS Barrett, B. and Rajandream, M. A.

TITLE Direct Submission

JOURNAL

COMMENT

Submitted (12-May-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrel@sanger.ac.uk

Notes: All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry SC9920, accession no. Z48639 and at the end by cosmid 8021.

FEATURES

source

1. 29987
location/Qualifiers

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 8156"

/map="13R"

/complement(<1..356)

/note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YM9920.13c"

/codon_start=1

/product="unknown"

/protein_id="CA89242.1"

/db_xref="GI:809082"

/db_xref="SWISS-PROT:O03496"

/translation="MTDVFSSOEIELEFVKEFLIANNPANNKINNENKDAVLTQIEHD FRYLYIKDGLPMLNESTRLLIPPTFISCLLRSHQILASKIDQSEFLSAVKDELTL EANNITFEVLEDFWA"

misc_feature

1. 140
/note="overlap with SC9920 Z48639 S. cerevisiae chromosome XIII cosmid 9920"

gene

/complement(851..1312)

CDS

/gene="TIF11"

/complement(851..1312)

/gene="TIF11"

/note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, initiation factor 1a"

/codon_start=1

/product="TIF1p"

/protein_id="CA89243.1"

/db_xref="GI:809083"

/db_xref="SWISS-PROT:P38912"

/translation="MGKNTKGGKRGKRGKNDGPKRELLYKEGGEVATITKMLGN GRVASCDFGKRAHIRGLKRYVMGOGDILVSLDFODDQDVVHKYVLEART LKNGELPENKINETDINFGESEDDVNFEGNADDEDEGEDELIDDI"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-crenolose-phosphate synthase and PIR:S27471; PS00213 lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIVASFLPTPOEDVYNSDPAKIVSSMIKVCNNNELS NNKGRSSSVTSASHYITGLPQEAQINQEPQDRAVAVGSPATGVNTNHEMELSSQFL BELTANATHAANSIGIPANNPVSGSTIQRSPVEFFFSAPARVCSPOEASASISA SRSRHHNDLSSLMKNPNLSEDSHPRVRSSGSAVITPVKSVDPDPAVDVAKY REERQOASLPSPMKRSGSTAGDSISASSSNLRYSOOFQDNFIEDTSDSDIDLE TDAIKYVVPKPGYSSNAKILASIMRBSYELFKLPTIYDSKGNSSKRAVNIAY AKYIVKEPVSMTGVTGIPITDELPHVECHKISKLDQDFSSFPVYVTDITFKATINIA

misc_feature

/note="PS00213 lipocalin signature"

CDS

/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:YTP3_YEAST P38430"

CDS

/note="YM8156.05, unknown, len: 301, CAI: 0.12, SW:YTP1_YEAST P38429 previously noted ORF"

CDS

/note="YM8156.06, unknown, len: 461, CAI: 0.11"

CDS

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.09, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.10, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.11, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.12, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.13, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.14, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.15, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.16, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.17, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.18, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.19, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.20, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.21, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.22, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.23, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.24, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.25, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.26, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.27, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.28, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.29, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.30, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.31, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.32, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.33, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.34, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.35, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.36, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.37, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.38, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.39, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.40, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.41, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.42, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.43, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.44, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.45, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.46, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.47, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.48, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.49, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.50, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.51, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.52, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.53, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.54, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.55, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.56, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.57, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.58, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.59, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.60, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.61, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

CDS

/note="YM8156.62, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

```

13243..14117
/gene="ipp2"
13245..14117
/gene="ipp2"
/note="PW8156.09, IPP2 gene (PPA2), len: 310, CAI: 0.14,
SM: PYZ_YEAST, P28239, inorganic pyrophosphatase,
mitochondrial; contains P500387 Inorganic pyrophosphatase
signature"
/codon_start=1
/product="ipp2p"
/protein_id="CA89250.1"
/db_xref="GI:809090"
/db_xref="SWISS-PROT:P28239"
/translation="NLLRNMLTSAKSIERLKQTLLTISNNHROFTIOGSKYT
GKKRYTLILNDEVSSFDVLDLNEHKRTVMIVVEPWTGKFEISNELPFNIV

```

QY	1	tgaaactgttgaatgaacaaagaagctctacaagaactcaaaagcttggtaacgatattcca	60
Db	27785	TGAAGTGTGATGAGCAAGAAGAACCTACACAGCTTCAAAGCTTGTGACGATTTCCA	27726
QY	61	ttcaaaaattcgtgaattgaaaccgaattcaacaagaatttccttgaaatnaatygatt	120
Db	27725	TTCAAAAATTTGTGATATGAAACCAAGCTTCACACAGATTTTCTTGATTTAAATGAGTT	27665
QY	121	ataataatgaaatgaaggagaccgacgagccaaagaagaattacaagaatttaataac	180
Db	27665	ATATATATGATATAGGGGAGACCGCTTAGGCCCAAGAACTGAAATTTAAGATTTTATTAC	27606
QY	181	ttcacagctggttcaacgatatcgaagaacattgcgaacttgaaggttagtgttccaactt	240
Db	27605	TTCAACAGCTGTTTAAACGATATCGAAGACATTGTGCACCTTGAAAGTTAGTTGCACAACTT	27545
QY	241	atccacaactttttaaaaaaacgctcagctaaccttaaaacaacatcaagaacgcgcatctc	300
Db	27545	ATCCACAATTTTAAAAAAACAGCTCAGTACCTTAATAACAACCTTAAGCACGCGCTTTATC	27486
QY	301	tgaagatctcaatatcgcacacaggttaacggttagtctcatcgtcgcgcagttcaagcgtaatt	360
Db	27485	TGAGGATTCAAATATGCGACACAGGTAAGGTATGTTCAACGTGCGCCAACTTCAGGCTCATTT	27426
QY	361	aacgcgcagttaaccttctccatccaaagcagttatcttttacccttagcgttagctcgtcac	420
Db	27425	AACGCCAGTAACCTTCTTCATCCAAAGAGTATTATTTTACTACAGCGGTACTCGTCTAC	27366
QY	421	ttcccttgaaatttcaagaccagatggtctcataaatlbggttaggattgctccttcaagta	480
Db	27365	TTCCCTGAAATTTAAGACCAAGATGTTTCATTTAAAGGGTTAGGATTCGCTCTTTACATA	27306

QY	461	caaaagaacataatgtggaatttggaaatttaataagaacataaaggaaacttaattcc	540
Db	27305	CAAAAGAACATTAATATGCAACTTGGAAATTAATAAGACATTAAAGAAACCTTAATTTC	27244
QY	541	aagtttcaaaagcccaatgttgaagtttataatgcgttcgaagaaatgaatcaatttga	600
Db	27245	AAGTTTCAAAAGCTGCCATATGTGTGAGGTTTTATTGCCGTTCCAGTATATATTAATTGA	27185
QY	601	aaacactcttgcgttagaagaatgatataccctaattctgttaaggcaagtacaaaata	660
Db	27185	AAACCACTTGCGGAGGAGATGATATATCCATTCTGTTAGGCAAGTGCAAAATA	27125
QY	661	aaaaaaccttggaaaaaattcgttactttcttatagatagatatagatagttt	720
Db	27125	AAAAAACCTTGGAAAAAATTCGTTACTTTCTTATGATATTAATATATATATGCTTT	27065
QY	721	gcttaataatgaagttattatcgcgttccttgatattccctaatttaataattcctt	780
Db	27065	GCTTATATATGAGGTATTTATCGGCTCTTGATTCCTCATTTTATTAATAAATCTTT	27005
QY	781	taaaatgcaatttcctgctccttcttgttcttgcctcgtatatttttttttttggacaacg	840
Db	27005	TAAATGCAATTTTCGTCGCTCTTGTGTGTCGTGATTTTTTTTTTTTGGACACAG	26944
QY	841	gatggaaaaaccttgaatgatttacttacttatttaagtactaaaataatcgagatt	900
Db	26945	GATGGAACAACTTGATATTTTATTTACTCTTATTTTAAGTACAAAATATCGAGATTT	26885
QY	901	cagagacaaaacataagaatttcttcttgcagaaaaaataaaacgaaataatgttgtct	960
Db	26885	CAGGACAAAACATATGAATTTTCTTGTGCAGAAAAATTAACGAAATTAATGAGCTT	26825
QY	961	tgaactacgactgtctgcataagagaaacgaacgaatgctacaaattctcaacatt	1020
Db	26825	TGACTACGACTGTGTGCATATGAGAAACCAAGAACGAATGCTACAAATTTCTCAACATT	26765
QY	1021	tctagtctcctatacatctccatactcctatgttgcctacgctatatacgcaacttcaacc	1080
Db	26765	TCTAGTCTCCTATACATCTCCATCTCCATTCCTATGTGTACCGCTATATGCAACTTCCAAACC	26705
QY	1081	agaagggcctaagaagtaatcccgaaagataaactctgcatacgcaaaactcaact	1140
Db	26705	AGAAAGGCTTAAGCAGATTAATCTCGAAGCTTAATCTCCAGCAAAAACCTTAACAT	26645
QY	1141	tatgctaaattcccaacccttcttgtgcctttcttcaactctcaatctccagtaccac	1200
Db	26645	TATGCTAATTTCCAACTTTTGTGTGCTTTTTTAATCTCAATTAATCTAGTAACAC	26585
QY	1201	ttcacataaagttccaagacgaattttagagcttaagtataaccgaagtataacg	1260
Db	26585	TTCACTAATTAAGTTTCAAGACGCAATTTTATAGGCTTAGGTATTAATCCACAGGTTATTACGC	26525
QY	1261	tgcatttgcaaaaccttggcaattcagccaagttctgtgaagaacttaacgaatagtgttc	1320
Db	26525	TGCATTGCAAAACCTTGCGCAATTCAGCAGTTCGTGAAGACTTAACGAATAATGTTGCTGC	26465
QY	1321	gatgctatctgacctatattcttgaaacccgttttaagatttgattataatcatatttaa	1380
Db	26465	GATGCTATTCGACTTAATTTGTGGACCGGTTTAAATTTGATTAATATCATTTATTAA	26405
QY	1381	tccaaagagctctatactctgaagattttacacatgaattctcgtaaatlatttggagttcag	1440
Db	26405	TCCAAAGAGCTCTATCTAGTGAAGATTTTACCATGAATTCGATATTTTGGAGTTTCAG	26344
QY	1441	gaattttatttgcacaaataaactggaagaatatatttaaccgttaaatgctttgactac	1500
Db	26344	GAAATTTATTTTGCACCAATPAACGAAGAAATATTTTACAGGTAAAGCTTTTACATAC	26284
QY	1501	gtactttaaacttaataccgacttgcacaataagctacaacagttatatttggcaaccac	1560
Db	26285	GTACTTAACCTTAATACCGCAATTCGCACATACGATATACACAGTTATTTTGGCAACGATC	26225
QY	1561	gcttttttttggacttgcgcagcaacacatgcttatagagcaattacagaagaagcttcat	1620

Db	25145	CCGGTAGGAACCTTGATCGGAGGGGTAACTCAGGAGATAGAAATCCTCCAAATAAGACTAA	25086
Qy	2701	caagtaactacttaagtatataaaaaaggtgttctactctccatccatccagaataatlaact	2760
Db	25085	CAAGTATACCTTATGGGATATAAAAAGGTGTACCTCTCCATCCAGGAAATTAATGCT	25026
Qy	2761	aacglatcaaaagaaaactctla tlyggttccgcgccaatcaacaccttaagtaagcct	2820
Db	25025	AACGATTAATGAAGAAAACCTTATATGTGGTCCCGCAATCAACCCCTTAAGCTTAAGCT	24966
Qy	2821	gatatctccagaagcttgtacaagatctcttaaaaaatacacaactaagcgacaatgt	2880
Db	24965	GATATTTCCTAGAGCCTTGTAAGATACCTTACAAAATATACAACTAAGCACAATGCT	24906
Qy	2881	gaagataatgtatgggaatagcaatgaaaaataacgatalttgagataatlygggaggaataa	2940
Db	24905	GAAGTATATGATGGGAATAGCAATGAATAATACGATTTGAGGATTAATGGGAGAGATAAA	24846
Qy	2941	gaatcacca 2948	
Db	24845	GAAATACA 24838	

RESULT	2	LOCUS	D50276	4156 bp	DNA	PLN	10-FEB-1999
DEFINITION		Saccharomyces cerevisiae Hst1 gene, complete cds.					
ACCESSION		D50276					
VERSION		D50276.1					
KEYWORDS		Hst1; high-copy suppressor of TFP sensitivity; Saccharomyces cerevisiae (strain:RAY-3A) DNA.					
SOURCE		Saccharomyces cerevisiae					
ORGANISM		Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.					
REFERENCE		1 (bases 1 to 4156)					
AUTHORS		Tsuchiya,E.					
TITLE		Direct Submission					
JOURNAL		Submitted (13-APR-1995) to the DDBJ/EMBL/Genbank databases. Eiko Tsuchiya, Hiroshima University Faculty of Engineering, Department of Fermentation Technology, Kagami-yama, Higashi-Hiroshima, Hiroshima 739, Japan (E-mail:etsuchi@ipc.hiroshima-u.ac.jp, Tel:0824-24-7868, Fax:0824-24-7868).					
REFERENCE		2 (bases 1 to 4156)					
AUTHORS		Tsuchiya,E., Matsuzaki,G., Tsukao,A. and Miyakawa,T.					
TITLE		Molecular cloning and characterization of S. cerevisiae Hst1 gene					
JOURNAL		unpublished (1995)					
REFERENCE		3 (sites)					
AUTHORS		Tsuchiya,E., Matsuzaki,G., Kurano,K., Fukuchi,T., Tsukao,A. and Miyakawa,T.					
TITLE		The saccharomyces cerevisiae Ssd1 gene is involved in the tolerance to high concentration of Ca2+ with the participation of Hst1/HRC1/BF1					
JOURNAL		Gene 176 (1-2), 35-38 (1996)					
FEATURES		97075905					
JOURNAL		Location/Qualifiers					
MEDLINE		1. 4156					
FEATURES		/organism="Saccharomyces cerevisiae"					
SOURCE		/strain="RAY-3A"					
		/db_xref="taxon:4932"					
gene		946..3693					
CDs		/gene="Hst1"					
		946..3693					
		/gene="Hst1"					
		/standard_name="high-copy suppressor of TFP sensitivity"					
		/codon_start=1					
		/protein_id="BAA08819.1"					
		/db_xref="GI:1816450"					
		/translation="MSNRNDSMLRTTSDKAIASQDKRKSEVLLIAQSIDNEIRSV					
		KNLRKISGMDLIDPDLDFGEGSGRSGSGTSSASMSPTTYNRTYSDP					
		TLPLNRHGSRGSGTSSNKRKGVGLGKKGVHSPKRLNAYLKNLWPAQHPR					
		VKPNFLEIVDDIQAQNLDSNGEDNGNSNNDDKGSASSTVAKYVERLSIKERY					
		NRGSRHGNASLIRPSTLRSTLYEFDNEDDCKGSDKQPOPECHYDGDGIGEST					
		SLRTTELELTISNSAGLTNDATLTARTLSMASSYSDSKQPOPECHYDGDGIGEST					

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825

Sequence: 1 acctaccctttttctatct.....aaaataacgattaacatt 1825

Scoring table: IDENTITY_NTC
Gapop 10.0 ; Gapext 1.0

Searched: 1472140 segs, 824899755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_ov:*
- 21: em_pat:*
- 22: em_ph:*
- 23: em_pl:*
- 24: em_ro:*
- 25: em_sts:*
- 26: em_sy:*
- 27: em_un:*
- 28: em_vl:*
- 29: em_vl:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1823.4	99.9	1825	8	SCYR117W	249617 S.cerevisiae
2	1553.4	85.1	1706	8	SCU77137	U77137 Saccharomyc
3	342	18.7	1555	8	SCYR116W	249616 S.cerevisia
4	195	10.7	99360	8	NCH11N2	AL513444 Neurospor
5	187	10.2	36920	8	SPAC3H1	268144 S. pombe ch
6	137.2	7.5	1275	8	AF353722	AF353722 Arabidops
7	136	7.5	686	6	A86270	A86270 Sequence 92
8	136	7.5	686	6	ARI55763	ARI55763 Sequence
9	136	7.5	686	6	E66288	E66288 Genome DNA
10	121	6.6	935	8	SCYR118C	249618 S.cerevisia
11	117.4	6.4	1859	9	AF064867	AF064867 Homo sapi
12	117.4	6.4	2963	9	AB016068	AB016068 Homo sapi
13	117.4	6.4	2966	9	HSY13834	Y13834 Homo sapien
14	117.4	6.4	2968	6	E32056	E32056 Human AFCL
15	114.2	6.3	1340	3	AF358443	AF358443 Physarum
16	98.4	5.4	78592	3	AC004335	AC004335 Drosophila
17	98.4	5.4	102413	2	AC020024	AC020024 Drosophila
18	92.8	5.1	41327	3	CBRG47J19	CBRG47J19 Caenorhab
19	91.6	5.0	40767	3	CRC04F12	281461 Caenorhabd
20	88	4.8	317511	1	CJ11168X3	AL139076 Campyloba
21	86.6	4.7	582	8	CNS018P	AL110644 Botrytis
22	81.6	4.5	660	8	CNS01AH3	AL112927 Botrytis
23	72.2	4.0	2161	9	AK027874	AK027874 Homo sapi
24	66.8	3.7	1072	6	AX055334	AX055334 Sequence
25	66.6	3.6	124808	2	AP000652	AP000652 Homo sapi
26	65.6	3.6	16488	1	AE004992	AE004992 Halobacte
27	65	3.6	256172	2	AC005139	AC005139 Plasmodiu
28	64.4	3.5	169423	2	AC069309	AC069309 Mus muscu
29	64.4	3.5	186306	2	AC083946	AC083946 Mus muscu
30	63.6	3.5	1245	6	AX055202	AX055202 Sequence
31	63.6	3.5	19715	1	AE000555	AE000555 Helicobac
32	63.6	3.5	234112	3	PFMAL1P2	AL035475 Plasmodiu
33	62.4	3.4	13583	3	AE001374	AE001374 Plasmodiu
34	61.8	3.4	98734	2	PFMAL1P2	AL031745 Plasmodiu
35	61.6	3.4	14943	1	AE001528	AE001528 Helicobac
36	59.8	3.3	12029	3	AE001392	AE001392 Plasmodiu
37	59.8	3.3	110000	2	AC087899	AC087899 Mus muscu
38	59.2	3.2	114736	3	PFMAL1P3	298547 Plasmodiu
39	59.2	3.2	253505	3	PFMAL1P7	AL034559 Plasmodiu
40	58.8	3.2	98899	2	AL603743	AL603743 Danio rer
41	58.8	3.2	318221	2	PFMAL1P3	AL049184 Plasmodiu
42	58.8	3.2	318503	2	AC074166	AC074166 Mus muscu
43	58.2	3.2	172239	2	AC069587	AC069587 Homo sapi
44	58.2	3.2	12029	3	AE001412	AE001412 Plasmodiu
45	58.2	3.2	187469	9	AL357312	AL357312 Human DNA

ALIGNMENTS

RESULT 1
SCYR117W 1825 bp DNA PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome X reading frame ORF YCR117W.
DEFINITION
ACCESSION Z49617.1 Y13136
VERSION Z49617.1 GI:1015836
KEYWORDS
SOURCE
ORGANISM

baker's yeast.
Saccharomyces cerevisiae

Eukaryotes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 1825)
AUTHORS Rose, M., Koetter, P. and Entian, K.D.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1825)

AUTHORS MIPS.
TITLE Direct Submission

Submitted (25-SEP-1995) Data collected by MIPS on behalf of the
European yeast chromosome X sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

QY	331	aggcttattcaltgltttrattcattgaagcgnlttcgcacactccctaataatccgltgaaa	390
Db	61	AGGCTTTTATATCATGTTTGATCTTATAGACGATTTCTCGACCAATCTCAATATTCCTGTGGAAA	120
QY	391	ttatcaatttcgggtgttcgcgaatgcaccaatttccttcgaaacttaactgaactagaca	450
Db	121	TTAAATATTTCtGGGTTCTCGATTGGCCCAATTTCTTTCGAATCTTACTTACGTACAGA	180
QY	451	caataccagaagctatctgaaacaaagtgtgcacactgtgcctggaagaaatgtgat	510
Db	181	CAGTACACAGAGCTATCTGAACCAAGATGGCCACTGTGCTGGAAGACGAATTTGATGAT	240
QY	511	gaaactttcctaataatcaagaactactccggcgccaaagcccaagttccatttgcgt	570
Db	241	GAACCTTTTCATTAACCAAGAACTACTCCGGGCAAGGCCAATTTCTCATTTTGGGT	300
QY	571	gaagctataaacttagcccaaaagctagttttcatcaaatcgaacttcccctaaatc	630
Db	301	GAGGCTATTAACCTACCCCAAAAGCTAGTTTTCATCAATACAGACTTTCCTCAAAATC	360
QY	631	tggcacatggcggtttcccttattgaaatgcagtcgcctgcgaactagaattcatalgtctcc	690
Db	361	TGGACATGGCGGTTCTTATTGAAATGACAGTCGCGCCAGCTAGATTCAATGTGCTCC	420
QY	691	actgcgcacagaagttatgtctcttggtctctatccagttgtctactgtgtgat	750
Db	421	ACTGTGCGACAGAGTTTAAAGCTTCTTGCGGTCTCTTATCCAGTTTGTCTACCTGGTTGAT	480
QY	751	ttagcaactctcttacctatgacatttttgcgcggagaagaaaaattggtttcaataatg	810
Db	481	TTGGCACTCTCTTACTATATAGCACTTTTGTCCTGGAGAAATAATTTGGTTCAATTAATTG	540
QY	811	accgtccaactatgatacacaagatgatcatcaagagctgcacttggcgtatgcattggt	870
Db	541	ACCGTCCAATATATGATATCACCATATATATGATCAAGAAGCTGTACTTTGGCTATGCTATGTGT	600
QY	871	ggcccaactccttacctgttcccttaagaactcttgaataatccctactgattcccttgg	930
Db	601	GGCCCAATCTTACTACCTGTTCTTAAAGTCTTGATTAATTCCTACATGATTTCCATTGCTTGG	660
QY	931	taactatgctctcttctgttctgttcgcaaatcttagcactgaaataatctccagcttc	990
Db	661	TACTATTATGGTCTCTTCTGTCTGTCTGTCCAAATCTTAAGCATATACATATTCACATCTTC	720
QY	991	atcatgcacatgttctaataagtctactccacttggagagcggatgaaataactat	1050
Db	721	ATCATGCCCATGTTTATATATAGTTCACTCATTTGGAGNAGCGTGAACATAAAAAATCTATT	780
QY	1051	gaaagtgtgcgatagagtttgggttccctctagataaagattttgtcaatgtacggtcca	1110
Db	781	GAAAGTTTGGCCATATGAGAGTTGGGTGCTCCCTAGATTAAGATTTTGTGCTATGACGGCTCA	840
QY	1111	aaaaagattccttcatccaacagataattccaaggtttgcacttcaactccaagagaatt	1170
Db	841	AAAAGATCTTCTCATTTCAAAACGGATTTTTCACAGGTTGGCCATTCCTCCCAAGAGATT	900
QY	1171	gtttgtgtcgaacctttagtgaacagtaattctacgtatgaaatlaagcgtgtttggcc	1230
Db	901	GTTTGTGTGACACTTTAGTGAACATATTTCTACTGATGAATAATTACGGCTGTTTTGGCC	960
QY	1231	catgaatcgtgtcactgtgcacaaaaacacacatcgttaataatgtatcatcttaagtaatg	1290
Db	961	CATGAATATCGGTACATGGCCAAAATAAACACATCGTTAATATATGTCATCTTTAATGCAATTG	1020
QY	1291	caacacctctcatatttctccctttccacagacactcaagaataataatcatttaaacac	1350
Db	1021	CACACCTTCTCTCATTTTCTCCCTTTTCACACACATCTACAGAAATACATCATTTTACAC	1080
QY	1351	acctctggcttttctcttaagaagaagctcactgtgcgatttgtgtatccggtttacaaga	1410
Db	1081	ACCTTGCGGCTTTTCTTAAGAAAGTCCACTGCGAGTTTGTGTGATCCCGTTATCTCAAG	1140
QY	1411	gaattcccaatcatcttgattatgtatttaaacagtatttaactccaactgaaagt	1470

THIS PAGE BLANK (13PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds
(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1
Perfect score: 1825
Sequence: 1 acctaccctttttttctatctc.....aaaataaagatttaaacatt 1825

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pt.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_com.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_hgt.*
31: em_hgt_hum.*
32: em_hgt_inv.*
33: em_hgt_hum.*
34: em_hgt_inv.*
35: em_hgt_inv.*
36: em_hgt_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1823.4	99.9	1825	8 SCUR117W	249617 S.cerevisiae
2	1553.4	85.1	1706	8 SCUR7137	077137 Saccharomyces
3	342	18.7	1555	8 SCUR116W	249616 S.cerevisiae
4	195	10.7	99360	8 NCBI11N2	AL131444 Neurospor
5	187	10.2	36920	8 SPAC3H1	268143 S. pombe ch
6	137.2	7.5	1275	8 AF353722	AF353722 Arabidops
7	136	7.5	686	6 AB6270	AB6270 Sequence 92
8	136	7.5	686	6 AR155763	AR155763 Sequence
9	136	7.5	686	6 E66288	E66288 Genome DNA
10	121	6.6	935	8 SCUR118C	249618 S.cerevisiae
11	117.4	6.4	1859	9 AF064867	AF064867 Homo sapi
12	117.4	6.4	2963	9 AB016068	AB016068 Homo sapi
13	117.4	6.4	2966	9 HS138334	Y138334 Homo sapien
14	117.4	6.4	2968	6 E37056	E37056 Human AF1
15	114.2	6.3	1240	3 AF358443	AF358443 Physarum
16	98.4	5.4	78592	3 AC004335	AC004335 Drosophila
17	98.4	5.4	102413	2 AC020024	AC020024 Drosophila
18	92.8	5.1	41327	3 CBRG47J19	AC084676 Caenorhab
19	91.6	5.0	40767	3 CEC04F12	281464 Caenorhabd
20	88	4.8	317511	1 C111168X3	AL139076 Campyloba
21	86.6	4.7	582	8 CNS018P	AL110644 Botrytis
22	81.6	4.5	660	8 CNS018P3	AL112927 Botrytis
23	72.2	4.0	2161	9 AK027874	AK027874 Homo sapi
24	66.8	3.7	1072	6 AX055334	AX055334 Sequence
25	66.6	3.6	124808	1 AP000652	AP000652 Homo sapi
26	65.6	3.6	16488	1 AE004992	AE004992 Halodactyl
27	65	3.6	236172	2 AC005139	AC005139 Plasmodiu
28	64.4	3.5	169423	2 AC069309	AC069309 Mus muscu
29	64.4	3.5	186306	2 AC083946	AC083946 Mus muscu
30	63.6	3.5	1245	6 AX055202	AX055202 Sequence
31	63.6	3.5	19715	1 AE000555	AE000555 Helicobac
32	63.6	3.5	234112	3 PFMA14P2	AL035475 Plasmodiu
33	62.4	3.4	13383	3 AE001374	AE001374 Plasmodiu
34	61.8	3.4	98734	2 PFMA11P2	AL031745 Plasmodiu
35	61.6	3.4	14943	1 AE001528	AE001528 Helicobac
36	59.8	3.3	12029	3 AE001392	AE001392 Plasmodiu
37	59.8	3.3	110000	2 AC087899_0	AC087899 Mus muscu
38	59.2	3.2	114736	3 PFMA13P3	298547 Plasmodiu
39	59.2	3.2	253305	3 PFMA13P7	AL034559 Plasmodiu
40	58.8	3.2	98899	2 AL603743	AL603743 Danto rer
41	58.8	3.2	318221	2 PFMA13P3	AL049184 Plasmodiu
42	58.8	3.2	318503	2 AC074166	AC074166 Mus muscu
43	58.4	3.2	172239	2 AC069587	AC069587 Homo sapi
44	58.2	3.2	12029	3 AE001412	AE001412 Plasmodiu
45	58.2	3.2	187469	9 AL357312	AL357312 Human DNA

ALIGNMENTS

RESULT 1
SCUR117W 1825 bp DNA PLN 11-NOV-1997
LOCUS S.cerevisiae chromosome X reading frame ORF YXN117W.
DEFINITION 249617 Y13136
ACCESSION Z49617.1 GI:1015836
VERSION 249617.1 GI:1015836
KEYWORDS
SOURCE
ORGANISM

Saccharomyces cerevisiae

baker's yeast.

Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

REFERENCE 1 (bases 1 to 1825)

AUTHORS Rose, M., Koetter, P. and Entian, K. D.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1825)

AUTHORS MIPS.

TITLE Direct Submission

JOURNAL Submitted (25-SEP-1995) Data collected by MIPS on behalf of the

European Yeast Chromosome X sequencing project. MIPS at the

Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152

QY 271 aaaccactctataatcaacaggaagaagaaaaaaaggaataagaaaaactgc 330
|||||
Db 1 AAACCACTCTCTAATTAACAGGAAAGGAGGAAAAAAGAGGAAATAGCAAACTGC 60

Qy	331	agccttctatcattgttgaattcttaagacgattctcgcacactcctaataatcccgltgaaa	390
Db	61	aggcctttatctacgttttatsattttaagacagattctcgaccattcccaatcccggtgaaa	120
Qy	391	ttaacattctcgtgttcgcgaattgcaccaatttcttcgcgaactcttaactgaagtaga	450
Db	121	tttaattcttttcggtgttcgcgaattgcaccaatttcttcgcgaactcttaactgaagtaga	180
Qy	451	cagttaccagaagctatctgaacaaagtctgcaccctgtgcgttgaagacgaattgatgat	510
Db	181	cagttaccagaagctatctgaacaaagtctgcaccctgtgcgttgaagacgaattgatgat	240
Qy	511	gaaactttcctaataacaaagaaactactccggcgccaaagccaaagttccatttcggt	570
Db	241	gaaacttttctatataacaaagaaactactccggcgccaaagccaaagttccatttcggt	300
Qy	571	gacgtcataaacttagcccaaaagcagtttttcaactcaaatcgacacttccccaataatc	630
Db	301	gacgtctatataaacttagcccaaaagcagtttttcaactcaaatcgacacttccccaataatc	360
Qy	631	tggacaatggtgcgtttctcttattgaatgtcagtcgcgcgcagtcagaattcatalgtctcc	690
Db	361	tggacaatggtgcgtttctcttattgaatgtcagtcgcgcgcagtcagaattcatalgtctcc	420
Qy	691	acgtgtcacagaagttatagctctctggctctcttaccagtttgtctaccctggtgat	750
Db	421	actgtcccaacagagtttaagctctctggctctcttaccagtttgtctaccctggtgat	480
Qy	751	ttgcacatctcttaactatagccatttgcgcggagaagaaaattggtttcaataaatg	810
Db	481	ttgcacatctcttactatagccatttgcgcggagaagaaaattggtttcaataaatg	540
Qy	811	accgtccaacatactgatacacaagatactgaacaaagctcgaacttgcgtacgtcattggt	870
Db	541	accgtccaacatactgatacacaagatactgaacaaagctcgaacttgcgtacgtcattggt	600
Qy	871	ggcccaactccttaaccgttcttaagatccttgaabaattccctactgatctcccttg	930
Db	601	ggcccaactccttaaccgttcttaagatccttgaabaattccctactgatctcccttg	660
Qy	931	tacattatgctcttctgttcgttgcgttcgaacttagacgaatgaataatccagctctc	990
Db	661	tacattatgctcttctgttcgttgcgttcgaacttagacgaatgaataatccagctctc	720
Qy	991	atcatctgccatgttttaataaagtctactccattcggagagcgtgaaactgaataatctat	1050
Db	721	atcatctgccatgttttaataaagtctactccattcggagagcgtgaaactgaataatctat	780
Qy	1051	gaaagtctggcgaatagagttggtgtccctctagataaagattttgtcaattgaagctca	1110
Db	781	gaaagtctggcgaatagagttggtgtccctctagataaagattttgtcaattgaagctca	840
Qy	1111	aaaagatctcttcaatcaaacgcatcttccaaggttgccttcaactccaaggaatt	1170
Db	841	aaaagatctcttcaatcaaacgcatcttccaaggttgccttcaactccaaggaatt	900
Qy	1171	gttttgttcgacactttagtgaacagttatcttaactgaatgaataatcagtcgttttgcc	1230
Db	901	gttttgttcgacactttagtgaacagttatcttaactgaatgaataatcagtcgttttgcc	960
Qy	1231	catgaaatcgttcaactcgtgcacaaaaaacacacatcgattataatgtcatcttaagccaattg	1290
Db	961	catgaaatcgttcaactcgtgcacaaaaaacacacatcgattataatgtcatcttaagccaattg	1020
Qy	1291	caaacctctctcaatttctccctttcacacagcatctaaagaataaactcaactttaaac	1350
Db	1021	caaacctctctcaatttctccctttcacacagcatctaaagaataaactcaactttaaac	1080
Qy	1351	acctctgcgttttctcttagaagaagctcactcgtgcagtttgcgtatcccgattacactaag	1410
Db	1081	acctctgcgttttctcttagaagaagctcactcgtgcagtttgcgtatcccgattacactaag	1140
Qy	1411	gaattcccatcatcatgattgattatgtaattaaacgatttaactccacgcgaatgt	1470

THIS PAGE BLANK (1 OF 10)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342
Sequence: 1 MEDUKTLDPNIPKILIS.....HPTLARSTALDVSEKKK 453

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2336	99.7	453	1 ST24_YEAST	P47154 saccharomyc
2	903	38.6	474	1 ST24_SCHPO	Q10071 schizosacch
3	698.5	29.8	475	1 FACT1_HUMAN	O75844 homo sapien
4	328	14.0	426	1 YHRN_BACSU	P40769 bacillus su
5	161	6.9	286	1 HTPX_PASMU	P57846 pasteurella
6	150.5	6.4	310	1 HTPX_HELPY	O92K54 helicobacte
7	150	6.4	335	1 HTPX_ARCFU	O30004 archaeoglob
8	147.5	6.3	283	1 HTPX_HAEIN	P48400 haemophilus
9	142.5	6.1	310	1 HTPX_HELPY	O25582 helicobacte
10	139	5.9	366	1 HTPX_AERPE	O9Y467 aeropyrum p
11	138.5	5.9	258	1 HTPX_METTH	O26669 metanobact
12	130.5	5.6	292	1 HTPX_PYRHO	O58997 pyrococcus
13	130.5	5.6	317	1 HTPX_TREAC	O9H1V2 thermoplas
14	127.5	5.4	2136	1 YCF2_MARPO	P09975 marchantia
15	120	5.1	292	1 HTPX_BUCAL	P57406 buchneria ap
16	119.5	5.1	2294	1 YCF2_ARATH	P56786 arabidopsis
17	119	5.1	297	1 HTPX_STRGC	O30795 streptococ
18	116	5.0	293	1 HTPX_ECOLI	P23894 escherichia
19	113	4.8	284	1 HTPX_METJA	O59076 methanococ
20	113	4.8	298	1 HTPX_BACSU	O31657 bacillus su
21	109	4.7	506	1 YCF2_EUGER	P31921 europa gra
22	109	4.7	595	1 YCF6_METJA	O58682 methanococ
23	107.5	4.6	1032	1 MT18_YEAST	P40469 saccharomyc
24	106.5	4.5	646	1 NTP1_HAEPV	O37319 heliothis a
25	106.5	4.5	1029	1 YCF5_YEAST	P43571 saccharomyc
26	105.5	4.5	873	1 COX1_ACACA	O37370 acanthamoeb
27	105	4.5	1427	1 SRB8_YEAST	P25648 saccharomyc
28	104.5	4.5	565	1 YHR9_YEAST	P38732 saccharomyc
29	104.5	4.5	1100	1 LOS1_YEAST	P33418 saccharomyc
30	104	4.4	302	1 HTPX_AOUAE	O67998 aquilex aeo
31	104	4.4	530	1 CP14_CHICK	P79760 gallus gall
32	103.5	4.4	185	1 GU01_RAT	P35994 rattus norv
33	103.5	4.4	450	1 VATA_SCHPO	O14265 schizosacch

ALIGNMENTS

RESULT	1	STANDARD	PRT	453 AA
ST24_YEAST	ST24_YEAST			
AC	P47154	1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC ENDOPEPTIDASE 1) (PPEP 1) (A-FACTOR CONVERTING ENZYME).			
DE	ST24 OR AF1 OR YHR17W OR J2032.			
GN	Saccharomyces cerevisiae (Baker's yeast).			
OS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_Taxid=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A. Entian K.D.;			
RA	Rose M., Koetter P., Entian K.D.;			
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A. AND FUNCTION.			
RA	Fujimura-Kamada K., Nouvet F.J., Michaelis S.;			
RL	"A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor."			
RN	J. Cell Biol. 136:271-285(1997).			
RP	[3]			
RA	FUNCTION.			
RL	MEDLINE=97218305; PubMed=9065405;			
RN	Boyartchuk V.L., Ashby M.N., Rine J.;			
RP	"Modulation of Ras and a-factor function by carboxyl-terminal proteolysis."			
RL	Science 275:1796-1800(1997).			
RN	[4]			
RP	FUNCTION.			
RA	MEDLINE=98365461; PubMed=9700155;			
RL	Tam A., Nouvet F.J., Fujimura-Kamada K., Stunt H., Sisodia S.S.,			
RN	Michaelis S.;			
RP	"Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAAX processing."			
RL	J. Cell Biol. 142:635-649(1998).			
RN	[5]			
RP	FUNCTION.			
RA	MEDLINE=98393572; PubMed=9725832;			
RL	Boyartchuk V.L., Rine J.;			
RN	"Roles of prenyl protein proteases in maturation of Saccharomycetes cerevisiae a-factor."			
RL	Genetics 150:95-101(1998).			
RN	[6]			
RP	FUNCTION.			
RA	MEDLINE=20158928; PubMed=10692417;			
RL	Schmidt W.K., Tam A., Michaelis S.;			
RN	"Reconstitution of the Ste24p-dependent N-terminal proteolytic step in yeast a-factor biogenesis."			
RL	J. Biol. Chem. 275:6227-6233(2000).			
RN	[7]			

34	103	4.4	645	1 VP74_NPYAC	P15963 autographa
35	103	4.4	1101	1 KOES_YEAST	O08217 saccharomyc
36	103	4.4	577	1 THT1_SCHPO	O09684 schizosacch
37	102	4.4	2493	1 YBA4_YEAST	P35194 saccharomyc
38	100.5	4.3	260	1 YMI6_RECAM	O21266 reclinomona
39	100.5	4.3	1040	1 RIK1_SCHPO	O10426 schizosacch
40	100	4.3	1325	1 MRP4_HUMAN	O15439 homo sapien
41	99.5	4.2	309	1 071AA_HUMAN	O76100 homo sapien
42	99	4.2	485	1 YC11_KLEPN	O48457 klebsiella
43	99	4.2	1024	1 RPOB_PLAFA	P21421 plasmodium
44	98.5	4.2	313	1 N02M_RHISA	O99817 thripcephal
45	98.5	4.2	417	1 O65A_DROME	P82982 drosophila

RA CHARACTERIZATION.
 RX MEDLINE=20285442; PubMed=10825201;
 RA Trueblood C.E., Boyartchuk V.L., Picologlou E.A., Rozema D.,
 RA Poultier C.D., Rhine J.,
 RT "The CAAX proteases, Atp1p and Rce1p, have overlapping but distinct
 RT substrate specificities.";
 RL Mol. Cell. Biol. 20:4381-4392(2000).
 RN [8]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=98409630; PubMed=9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.,
 RT "Endoplasmic reticulum membrane localization of Rce1p and Ste24p,
 RT yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal a-factor cleavage.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR MATING PHENOMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHENOMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on ways
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR	EMBL: Z49617; CAA89647.1; "	
DR	EMBL: U77137; AAB38271.1; "	
DR	MEROPS: M48.001: -	
DR	SGD: S0003878; STE24.	
DR	InterPro: IPR001915; Peptidase_M48.	
DR	InterPro: IPR000130; Zn_MTPeptide.	
DR	Pfam: PF01435; Peptidase_M48; 1.	
DR	PROSITE: PS00142; ZINC_PROPEASE; 1.	
KW	Hydrolase; Metalloprotease; Zinc; Transmembrane;	
FT	Endoplasmic reticulum; Pheromone response.	
FT	TRANSMEM	13
FT	TRANSMEM	90
FT	TRANSMEM	110
FT	TRANSMEM	122
FT	TRANSMEM	142
FT	TRANSMEM	168
FT	TRANSMEM	188
FT	TRANSMEM	198
FT	TRANSMEM	218
FT	TRANSMEM	307
FT	TRANSMEM	327
FT	TRANSMEM	358
FT	TRANSMEM	378
FT	METAL	297
FT	METAL	297
FT	ACR_SITE	298
FT	ACR_SITE	298
FT	METAL	301
FT	METAL	301
FT	METAL	390
FT	METAL	390
FT	ACT_SITE	394
FT	ACT_SITE	394
SO	SEQUENCE	453 AA; 52324 MW; 331CC9AED27C99DA CnC64; PROTON DONOR (BY SIMILARITY).

Query Match	99.7%	Score 2336	DB 1	Length 453
Best Local Similarity	99.8%	Pred. No. 3.7e-160		
Matches 452	Conservative	0	Mismatches 1	Indels 0
				Gaps 0

QY	1	MFDLTLIDHPNIPMKLIISGFSIAQFSFSESYLTROYOKLSETLPVLEBEIDDEFH	60
Db	1	MFDLTLIDHPNIPMKLIISGFSIAQFSFSESYLTROYOKLSETLPVLEBEIDDEFH	60
QY	61	KSRNRSRAKAKSFSTGCDVYNLAQKLVFTIKYDLPFRIMHMAVSLNAVLPRHMYTAAQ	120
Db	61	KSRNRSRAKAKSFSTGCDVYNLAQKLVFTIKYDLPFRIMHMAVSLNAVLPRHMYTAAQ	120
QY	121	SLCFGLGLLSLSTLVLDPLSLYSYSHPLVEEKEFGFNKLTYOLMTIDMKSLTAYAGGPTL	180
Db	121	SLCFGLGLLSLSTLVLDPLSLYSYSHPLVEEKEFGFNKLTYOLMTIDMKSLTAYAGGPTL	180

OY		181	YLFLKIDPKRPTDLAMVLEFVOVLAMIIVLFIIPMKNRPPLLEDGLKKSIESTA	240
Dd		181	YLFLKIDPKRPTDLAMVLEFVOVLAMIIVLFIIPMKNRPPLLEDGLKKSIESTA	240
OY		241	DRVGFPDRIKFVIDGSKRSSHNAFTGLPPTSRIVLFDTLVNSNSTDEITAVLAHEIG	300
Dd		241	DRVGFPDRIKFVIDGSKRSSHNAFTGLPPTSRIVLFDTLVNSNSTDEITAVLAHEIG	300
OY		301	HMQKNHYVMVFESQLHTFLIFSFLTSTYRTSTSYNPNFGFLENSTGSFVPVPIKEPEI	360
Dd		301	HMQKNHYVMVFESQLHTFLIFSFLTSTYRTSTSYNPNFGFLENSTGSFVPVPIKEPEI	360
OY		361	IIGFMELNDLTLPPECAMQPMVSLISPRHEHQAAVAKKLKGYKNLCALDLOIKNST	420
Dd		361	IIGFMELNDLTLPPECAMQPMVSLISPRHEHQAAVAKKLKGYKNLCALDLOIKNST	420
OY		421	MNVDPLYSSYHSHPILAERSTALDYSEKKNN	453
Dd		421	MNVDPLYSSYHSHPILAERSTALDYSEKKNN	453
<hr/>				
RESULT	2			
ST24_SCHPO	STANDARD:	PRT:	474 AA.	
ID AC	010071;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC ENDOPROTEASE 1) (PPSEP 1).			
GN	SPEC3HL.05.			
OS	Schizosaccharomyces pombe (fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
CC	Schizosaccharomycetales; Schizosaccharomycetaceae.			
OX	Schizosaccharomycetes.			
RN	NCBI_Taxid=4896;			
RP	(1)			
RC	SEQUENCE FROM N.A.			
RA	SPRAIN-972;			
RL	Gentiles S., Churcher C.M., Barrett B.G., Rajandream M.A., Walsh S.V.;			
CC	Submitted (DEC-1995) to the EMBL/GeneBank/DDJ databases.			
CC	-1- FUNCTION: PROTOTYPICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF FARNESYLATED PROTEINS (BY SIMILARITY).			
CC	-1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC METALLOPROTEASE).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb.sib.ch).			
CC	-----			
DR	EMBL: Z68144; GAA92258.1; .			
DR	MERP5: M48_001;			
DR	InterPro: IPRO01915; Peptidase_M48.			
DR	InterPro: IPRO00130; Zn_MTpeptidse.			
DR	Pfam: PF01435; Peptidase_M48; 1.			
KW	PROSITE: PS00142; ZINC_PROTEASE; 1.			
KW	Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane; Endoplasmic reticulum.			
FT	TRANSMEM	103	123	POTENTIAL.
FT	TRANSMEM	196	216	POTENTIAL.
FT	TRANSMEM	230	250	POTENTIAL.
FT	TRANSMEM	344	364	POTENTIAL.
FT	TRANSMEM	381	401	POTENTIAL.
FT	METAL.	332	332	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	333	333	BY SIMILARITY.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(Without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460a-3
Perfect score: 2948

Sequence: 1 tgaactgttgatgacacaag.....ggggagagataaagaatcaca 2948

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBank1:
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2948	100.0	29987	8	SC8156	249260 S.cerevisiae
2	1560.4	52.9	4156	8	D50276	D50276 Saccharomyces
3	1361	46.2	4130	8	YSCNCRIG	147821 Saccharomyces
4	1163	39.5	4225	8	SCU65682	065682 Saccharomyces
5	1065.4	36.3	3250	8	SCU32580	032580 Saccharomyces
6	931	31.6	3525	8	SCU63849	063849 Saccharomyces
7	786.8	26.7	3240	8	YSCBUL1	D50843 Yeast BUL1
8	703.8	23.9	3020	8	SCRDSTGEN	X88991 S.cerevisiae
9	725.2	7.6	270	8	YSCGRC5X	M87379 Yeast Eco R
10	143.8	4.9	27559	8	SC8339	243210 S.cerevisiae
11	75	2.5	2023	8	SCYNR068C	Z71683 S.cerevisiae
12	71.8	2.4	4775	8	SCU32938	U32938 Saccharomyces
13	68	2.3	780	11	CNS06K8B	AL402561 T7 end of
14	65.6	2.2	256172	2	AC005139	AC005139 Plasmodiu
15	65.6	2.2	310779	2	AC005140	AC005140 Plasmodiu
16	64.4	2.2	904	11	CNS06189	AL39967 T3 end of
17	64.4	2.2	86827	3	PFMAL3P5	AL034556 Plasmodiu
18	63.4	2.2	1920	3	PFMTB	X62393 P.falciparum
19	63.4	2.2	110000	2	PFMAL3P2_0	AL049185 Plasmodiu
20	62	2.1	253305	3	PFMAL3P7	AL034559 Plasmodiu
21	61.6	2.1	11441	6	AX083744	AX083744 Sequence
22	60.8	2.1	234112	3	PFMAL4P2	AL035475 Plasmodiu
23	60.6	2.1	202872	2	AC016160	AC016160 Homo sap1
24	60	2.0	5632	3	AF310892	AF310892 Dictyoste
25	59.4	2.0	3214	3	AF019980	AF019980 Dictyoste
26	59.4	2.0	318221	3	PFMAL3P3	AL049184 Plasmodiu
27	58.8	2.0	321003	2	PFMAL4P3	AL035476 Plasmodiu
28	58.4	2.0	78439	17	HST05197	U85197 Homo sapien
29	58.4	2.0	169546	2	AC004157	AC004157 Plasmodiu
30	58.4	2.0	245802	2	AC006279	AC006279 Plasmodiu
31	58.4	2.0	251124	3	HUAE000660	AE000660 Homo sap1
32	58.2	2.0	153098	3	PFMAL3P2	AL034558 Plasmodiu
33	58	2.0	168794	2	AC004688	AC004688 Plasmodiu
34	57.8	2.0	178137	9	AC011302	AC011302 Homo sap1
35	57.8	2.0	13684	3	AE001403	AE001403 Plasmodiu
36	57.8	2.0	153477	2	AC006278	AC006278 Plasmodiu
37	57.6	2.0	53932	2	AC023371	AC023371 Homo sap1
38	57	1.9	12029	3	AE001431	AE001431 Plasmodiu
39	57	1.9	163443	2	AC006280	AC006280 Plasmodiu
40	57	1.9	178273	2	AC005308	AC005308 Plasmodiu
41	57	1.9	196149	2	AC004709	AC004709 Plasmodiu
42	56.8	1.9	67970	3	PFMAL1P3	AL031746 Plasmodiu
43	56.8	1.9	149627	9	AC087428	AC087428 Homo sap1
44	56.6	1.9	92633	2	PFMAL4P1_3	Continuation (4 of
45	56.6	1.9	153477	2	AC006278	AC006278 Plasmodiu

ALIGNMENTS

RESULT 1
SC8156/c
LOCUS SC8156 29987 bp DNA
DEFINITION S.cerevisiae chromosome XIII cosmid 8156.
ACCESSION Z49260 Z71257
VERSION Z49260.1 GI:809081
KEYWORDS BUL1; cytochrome b5; DAG1; initiation factor 1A; inorganic pyrophosphatase; Ipp2; nitrate reductase; Orotate phosphoribosyltransferase; tau element; TIF1; TFS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP; URA10.
SOURCE Baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 29987)
AUTHORS Live-G. and Churcher,C.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29987)
AUTHORS Bartell,B. and Rajandream,M.A.
TITLE Direct Submission

THIS PAGE BLANK (1-3870)

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrel@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis. Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)
is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry
SC9920,
accession no. Z48639 and at the end by cosmid 8021.
location/Qualifiers

1. 29987
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8156"
/map="13R"

CDS

complement(1..356)
/note="YB8156.01c, unknown, len: 118, CAI: 0.14, overlaps
and extends YW9920.13c"
/codon_start=1
/product="unknown"
/protein_id="CA89242.1"
/db_xref="GI:809082"
/translation="MTDVFSSQSEIEFVKEFLIANPKNINNEKDAVLQIEHD
FRVIOYIKDGLPMLNESTRILPFDFTSICLRSHQIASKIDSQELFSAVKEQLLT
EANNITFEYVLDLFWA"

misc_feature

1. 140
/note="overlap with SC9920 Z48639 S.cerevisiae chromosome
XIII cosmid 9920"
complement(851..1312)
/gene="TIF11"
complement(851..1312)
/note="TIF11"
/note="YB8156.02c TIF11 gene, IF1A_YEAST, P3891,
initiation factor 1a"
/codon_start=1
/product="Tif1p"
/protein_id="CA89243.1"
/db_xref="GI:809083"
/db_xref="SWISS-PROT:P38912"
/translation="MGKNTKGGKRGKRGKNDSDGPKRELLYKEGGEVIAQITKLN
GRVASCPDGKRMARHIGKLRKYVMGQDILIVSLDPODDCDVYHKKINDLEART
LKNQELPENAKINETDNGFESDEVDVNEFGNDEDEDEDELDIDDI"

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YB8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST, P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESSMIKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADNGPILRANVGPATGVNHNEMKLSFOFL
BEITANATHAANGIPRANPYSSSGSTARQREVEFFAPASRVCSPQESASASSISA
SRSSAHNDLSSILKMPVLSFDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGSSIASSSNLRYSQOFODFIIDTDSDEDDIDLE
TDATKRYNPKGSGSNNAKLRSALSMRNSYELFKLPITVDSDRGNSLKNAAVIAV
AKTYKEPVSWGTGIPIDELPHEVCHRSIKKLEDDSSFPVVTDDITFGAYKXNA

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YB8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST, P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESSMIKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADNGPILRANVGPATGVNHNEMKLSFOFL
BEITANATHAANGIPRANPYSSSGSTARQREVEFFAPASRVCSPQESASASSISA
SRSSAHNDLSSILKMPVLSFDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGSSIASSSNLRYSQOFODFIIDTDSDEDDIDLE
TDATKRYNPKGSGSNNAKLRSALSMRNSYELFKLPITVDSDRGNSLKNAAVIAV
AKTYKEPVSWGTGIPIDELPHEVCHRSIKKLEDDSSFPVVTDDITFGAYKXNA

misc_feature

complement(2260..2301)
/gene="TPS3"
/note="PS00213 Lipocalin signature"
5199..6140
/note="YB8156.04, unknown, len: 313, CAI: 0.11, previously
noted ORF, SW:YTP3_YEAST, P38430"

CDS

complement(1..356)
/note="YB8156.01c, unknown, len: 118, CAI: 0.14, overlaps
and extends YW9920.13c"
/codon_start=1
/product="unknown"
/protein_id="CA89242.1"
/db_xref="GI:809082"
/translation="MTDVFSSQSEIEFVKEFLIANPKNINNEKDAVLQIEHD
FRVIOYIKDGLPMLNESTRILPFDFTSICLRSHQIASKIDSQELFSAVKEQLLT
EANNITFEYVLDLFWA"

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YB8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST, P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESSMIKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADNGPILRANVGPATGVNHNEMKLSFOFL
BEITANATHAANGIPRANPYSSSGSTARQREVEFFAPASRVCSPQESASASSISA
SRSSAHNDLSSILKMPVLSFDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGSSIASSSNLRYSQOFODFIIDTDSDEDDIDLE
TDATKRYNPKGSGSNNAKLRSALSMRNSYELFKLPITVDSDRGNSLKNAAVIAV
AKTYKEPVSWGTGIPIDELPHEVCHRSIKKLEDDSSFPVVTDDITFGAYKXNA

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YB8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST, P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESSMIKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADNGPILRANVGPATGVNHNEMKLSFOFL
BEITANATHAANGIPRANPYSSSGSTARQREVEFFAPASRVCSPQESASASSISA
SRSSAHNDLSSILKMPVLSFDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGSSIASSSNLRYSQOFODFIIDTDSDEDDIDLE
TDATKRYNPKGSGSNNAKLRSALSMRNSYELFKLPITVDSDRGNSLKNAAVIAV
AKTYKEPVSWGTGIPIDELPHEVCHRSIKKLEDDSSFPVVTDDITFGAYKXNA

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YB8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST, P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESSMIKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADNGPILRANVGPATGVNHNEMKLSFOFL
BEITANATHAANGIPRANPYSSSGSTARQREVEFFAPASRVCSPQESASASSISA
SRSSAHNDLSSILKMPVLSFDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGSSIASSSNLRYSQOFODFIIDTDSDEDDIDLE
TDATKRYNPKGSGSNNAKLRSALSMRNSYELFKLPITVDSDRGNSLKNAAVIAV
AKTYKEPVSWGTGIPIDELPHEVCHRSIKKLEDDSSFPVVTDDITFGAYKXNA

CDS

THIS PAGE BLANK

/note="PM2150.09; ip2p gene (ppa2), len: 310, CAI: 0.14,
 SW:IP21_YEAST.P28239, inorganic pyrophosphatase,
 mitochondrial, contains P500387 Inorganic pyrophosphatase
 signature"
 /codon_start=1
 /product="Ipp3p"
 /protein_id="CAAB9250.1"
 /db_xref="GI:809090"
 /db_xref="SWIS2-PROT:P28239"
 /translation="MLNRKNPAFLSKARSEIRLKQTLNLSIRNRHQRSTIQQGSKYTT
 GFKKYLILLNGEVSFHDVPLDLNHEKTYVMATVEYPRMTGTGFELSKELRNPFV

OY	481	caaaagacaccatcatggaacttggaaatttaataagaaccttaaggaaactttatccc	540
Dp	27305	CAACGAGACCTTATATGTGAACCTTGAAATTTAATAAGGACATTAGAACAACCTTTAATTC	27246
OY	541	aagttttaaacgcgcctatgttgtagttttatcgttcgtccgaaagttaaatgaattcga	600
Dp	27245	AAGTTTCAAAGCGCCCATGTGGTGTATTTATTTGCCTTGAGTAATGATTAAATTTGA	27166
OY	601	aaaccacttggcgtagcgaagattgataccccattctcgttaagcaagtgcacaata	660
Dp	27185	AAACCATTCTGGCCTAGGGAAGATGTGATCCCATTTCTGTGGCAAGACAAAAATA	27126
OY	661	aaaaaaccttggaaaanaattcgttctaatttcttatagatatgatatagttagtgttt	720
Dp	27125	AAAACACTTTGAAAAAATTCCTGTAATTTCTTAATAGAAATAATATATGATAGTTTT	27066
OY	721	gcttatagataaagllatttacgcgtcccttggatcccatctaatataaatctctt	780
Dp	27065	GCTTATAGATAAGGTATTTATTCGCTTCCTTGATATCCCATATATATAAATCTTT	27006
OY	781	taaaatgcattctcgtgtgccttttgttgcctcgtlatlcttttttttttttggacaactg	840
Dp	27005	TAAATGCATTTCTGGGCGCTTTTGATGCTCGATTTTTTTTTTTTTTGGACCAACG	26948
OY	841	gatggaaaaccttggatgattttatcccttattttaagttaaccaaatatcgagattc	900
Dp	26945	GATGGAACACCTTTGATGATTTTATATACCTTATTTTAGTTACTTAAATATCGAGATT	26886
OY	901	caggacaacaaactagaaattctcttctgcagaaaaaataaaacggaataaatttgact	960
Dp	26885	CAGGACAAACATATGAAATTTCTTTGTCAAGAAAAAATGAATTAATGATGCTT	26828
OY	961	tgactactgactgtctgcatagaagagaccaaagacaatgctacactctccaact	1020
Dp	26825	TGACTACTGACTGTCTGCATATAGAGAGAACCAAGAACATATGCAATTCATCAACTT	26766
OY	1021	tctagtctccatatcacatctccatactcctatgctctacogctatatgcaacttccaaoc	1080
Dp	26765	TCTAGTGTCTCTATACATCTCCATATCCTATGCTACCGTATATGCAACTTCCAAC	26706
OY	1081	agaagggtcctaagagatactcccgaaagattaactctgcgaagcaaaaacttaacat	1140
Dp	26705	AGAAAGGCTTAAAGAGATTAATCCTCGAACATTTAAATCTGCAATGCAAAAACTTAACAT	26648
OY	1141	tatgcatacttccaaaccttttttggtagcctttttttaaactcccatatcttagtaaac	1200
Dp	26645	TATGCTAATTTCCAACCTTTTGTGGTCCCTTTTAAATCTCAATATCTAGTACAC	26586
OY	1201	ttcacataatagttcgaagacgaacttttagcttaagttatccagttatccagttatcagc	1260
Dp	26585	TTCACATTAATTAAGTTTCAAGGACGATTTTATGGCTATAGGTATTAATCCACAGTTATTACG	26526
OY	1261	tgcaattgccaaaccttgcgaacttcagcccaattcgtagaagaacttaagaaattggttgc	1320
Dp	26525	TGCATTGCCAAACCTTGGCAATTCAGCCATTTGTGTAAGACTTAAAGAAATGTTGTGC	26466
OY	1321	gagtttatgaccttatattgtggaccggtttcagattttglatatatacatttataaa	1380
Dp	26465	GATGTTATTTGACCTTAATTTATTTGTGACCOGTTTATGATTTGTATATATCAATTAATAA	26406
OY	1381	tccaaagagcctatactattgaagatttttaccatgaattccctgaacatattvgagttcag	1440
Dp	26405	TCCAAAGAGCCTATACTTGAAGATTTTAAACATGAATTTCTGGAATATTTTGGAGTTTCAG	26346
OY	1441	gaatttatatttgcacaaataaacttgggaaatttttaccgcgcaatgccttggaccac	1500
Dp	26345	GAAATTTTATTTTCAACCAATATCTGAGGAATTTTATTCACGTCGAAGCTTTTGACATAC	26286
OY	1501	gtacttaaaccttaataccgacttcgcaactaagcatacaacagttaatttggcaacatc	1560
Dp	26285	GTACTTAACCTTAATACCGCAATTTGCAACTAAGCATATCAACAGTATTTTGGCAACATC	26226
OY	1561	gcttttttttggacttgcgacgacacacatgcttatatgagcaatlbaacagaaggtccat	1620

THIS PAGE BLANK (COPY)

Db 26225 GCTTTTGGACCTGCGACGACACCATGCTTATGACATTAAGAGAGCTCCAT 26166
 Oy 1621 gacactgtttccattctctgtcgaacacagcttccaaattttatacaaacacttttgg 1680
 Db 26165 GACAACTGTTTCTTCATCTGCTGACACATGCTTCAAAATTTTATACCAACACTTTTGG 26106
 Oy 1681 agggtaacaaattgtatctgtaagaacagcggaacctatggtcgtacatact 1740
 Db 26105 AGGGTTAAACCAATTTGTATTCGTAAGACAGCGGGACCTATGTCGCTCAATATCC 26046
 Oy 1741 gcatgccttctgcaatacatatggttctctgtccttcaagatgtaattacattcac 1800
 Db 26045 GCATGCCCTTTCATATTCATGCGGTTCTCGGCTTCAAGATGTAATTAATTAC 25986
 Oy 1801 agtaagtaacaaagaacgctgagacattccaaattgtctcaatctcggaataagtccta 1860
 Db 25985 AGTAGTAGACAGAAAGACGATGACCATTTCCAAATGTGTCCAATCTGGAAATAGTCTA 25926
 Oy 1861 cttegcactgtctgtccttctgtatataatccctgaagaataccttcaaacctcgttagg 1920
 Db 25925 CTTCGCACCTGCTGCTGCTTGGATTAATVCCCTGAAGATACCTTACAACTCTGTRAG 25866
 Oy 1921 aacctcgtgtatagaataaacctttagccttttactgtatatacctgttaaat 1980
 Db 25865 AACCTCGGTATATGAATTAACCTTTTACCTTTTACGTACTGTATACCGTTAAAT 25806
 Oy 1981 ttccatgtactataactcttttcaactactatattgaattctcgaagaccggtc 2040
 Db 25805 TTCCATGTACTATTAACCTTTTTCACATATATGGAATCTATCGACGACCGGCT 25746
 Oy 2041 ttgttacaagagagtgaaataacgagtttgggttctgttgaagaatttgaagac 2100
 Db 25745 TTGTGTACGAGAGATGAAAAATGAGTTTGGTGGTGAAGATTTGAGAGGAC 25666
 Oy 2101 tataagtaacctatactctgttatcgcgaatacaaacagtcgttcgtagtggat 2160
 Db 25685 TATAAGTACCTATTAATTTGATTTACGACTCAATACAGTCGTCGTCAGTGTAT 25626
 Oy 2161 tgaagtcgacatcacaagaagtagaagaagtgagcatcaataggttcgaagtttc 2220
 Db 25625 TGAAGTGTCAATCTAAGAGTAGAGAGAGGCGCATCAATAGCTTTCAGCTTTC 25566
 Oy 2221 ttttttaagtttcttactgtcctccatgaatttaagttcagtttgggtt 2280
 Db 25565 TTTTAAAGTTTATTTGCTCTCTCAATTAAGCTTGAAGTTTGGTGTGT 25506
 Oy 2281 ttgtgtgttacaattttcaattcaagaagaatttagctgtccttataatgccaat 2340
 Db 25505 TTTGTGGTTACATATTTTCAATTCACAAAGAGAAATTTAGCTGCTTTTAAATGTCAT 25446
 Oy 2341 agagataacgagacatgctcgttactacatacaagcgataagcgatcgtcgaag 2400
 Db 25445 AGAGATACGAGAGACATGCGCGTACATCAATCAAGCATTAAGCGATCGCTAGCAAG 25386
 Oy 2401 gataaacggaagtcgaagtttgaattgtcgaacagtccttgaacaatgtaacgagc 2460
 Db 25385 GATTAACGGAAGTCTGAAGTTTGAATGCTGACAGATCCCTTGAATGAATCCGACG 25326
 Oy 2461 gtaaaaacttaaaagattgtcattggttcgaattgatttactattgaccgaatta 2520
 Db 25325 GTAAAAAACCTAAAAAGTTTCTCATTTGGTCAATGATTTACTTATGATCCGAATTA 25266
 Oy 2521 gataaaatcgttggggaattagtgaggagacatcattgctcgtcgaacacgaat 2580
 Db 25265 GATTAATAATTCGGTGGGATCTAGTGGAGACATCATGTCGACACATCCAGT 25206
 Oy 2581 tctcgtcgaatgcgaatgacacacacccgttaataacaacgatatagagatccaaat 2640
 Db 25205 TCTCGCTCAATGCGAAGTACACACACCGTTAATACACAGATATAGGATCCAACT 25146
 Oy 2641 ccgctagaagacttgcagggagggttaactaaggatagaatcctccaataagactaa 2700

Db 25145 CCGTAGAGACTTGATGATGGAGGGGTAACTCAGGAGATGAATCTCCAAATTAAGCTAAA 25086
 Oy 2701 caaggtactacttaagatataaaagggttcaactcctccatcgaagaataatagct 2760
 Db 25085 CAAGGTAACTACTTATGATTAATAAAAGGTGTTCTCTCTCCATCCAGGAATTAATGCT 25026
 Oy 2761 aacgtattaagaanaacttattatcgtgttcccgcaatcaacaccccttaagtcct 2820
 Db 25025 AACGTATTAAGAAAACTTATTAATGAGGTCCCGCAATCAACCCCTTAAGCTTAAAGCT 24966
 Oy 2821 gataattcctagagctgttacaagatacatttacaataatacaactaagcaaatgct 2880
 Db 24965 GATTAATTCCTAGAGCTGTATACAGTACTTACAAAATTAACAACTAAGCGCAATGCT 24906
 Oy 2881 gaagataatgaggaatacgaacatacaagataatcgagataatgggaggaataa 2940
 Db 24905 GAAGATAATGATGGGAATAGCAATGAATAAATGAGATATGAGGAGATTA 24846
 Oy 2941 gaatacaca 2948
 Db 24845 GAATCACA 24838

RESULT 2

LOCUS D50276 4156 bp DNA PLN 10-FEB-1999
 DEFINITION Saccharomyces cerevisiae Hst1 gene, complete cds.
 ACCESSION D50276
 VERSION D50276.1 GI:1777313
 KEYWORDS HST1, high-copy suppressor of TFP sensitivity.
 SOURCE Saccharomyces cerevisiae (strain: RAY-2A) DNA.
 ORGANISM Saccharomyces cerevisiae
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE

1 (bases 1 to 4156)
 Tsuchiya, E.
 Direct Submission
 Submitted (13-APR-1995) to the DDBJ/EMBL/GenBank databases. EMBL
 accession number: F01111.1
 Tsuchiya, Hiroshima University Faculty of Engineering, Department
 of Fermentation Technology, Kagami-Yama, Higashi-Hiroshima,
 Hiroshima 739, Japan (E-mail: etsuchi@ipc.hiroshima-u.ac.jp,
 Tel: 0824-24-7868, Fax: 0824-24-7868).

REFERENCE

2 (bases 1 to 4156)
 Tsuchiya, E., Matsuzaki, G., Tsukao, A. and Miyakawa, T.
 Molecular cloning and characterization of S. cerevisiae HST1 gene
 Unpublished (1995)

REFERENCE

3 (sites)
 Tsuchiya, E., Matsuzaki, G., Kurano, K., Fukuchi, T., Tsukao, A. and
 Miyakawa, T.
 The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance
 to high concentration of Ca²⁺ with the participation of
 HST1/NRC1/BF1
 Gene 176 (1-2), 35-38 (1996)

TITLE

JOURNAL
 MEDLINE
 FEATURES
 source
 Location/Qualifiers
 1. 4156
 /organism="Saccharomyces cerevisiae"
 /strain="RAY-3A"
 /db_xref="taxon:4932"
 946..3693
 /gene="HST1"
 946..3693
 /gene="HST1"
 /standard_name="high-copy suppressor of TFP sensitivity"
 /codon_start=1
 /protein_id="BAA08819.1"
 /db_xref="GI:1816450"
 /translation="MSNRNDSMLPTSSDKAIAQSPDRKSEKVIILAAQSLDNEIRSY
 KNLRLISGMDLIDPELDIFGEGSSGRSRWSGTTSSASMPSTTTVNNTRYSDP
 PLENLEHGNGSIESNNKQNGNYLGIKKGVHSPRKLANAVLKNLWVPRANDHP
 VPDNLEHIVQNTIONLSGNDGNDGNSNNNDIEDNGEDKESQSYNNKNNNT
 NGLSRHGNASLIRRPSTRRSRYTERDDDDDDNGDSASEYNNKVEIRISIKERP
 SLDITTEILKISNSAGLTDNDATILARLISAGNSISDKKDPQPPGHGHDGDIQST

THIS PAGE BLANK (10/10)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825

Sequence: 1 acctaccctttttctatct.....aaaataacgattaacatt 1825

Scoring table: IDENTITY_NDC

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2344280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl :
1: gb_ba :
2: gb_hcg :
3: gb_in :
4: gb_cm :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vl :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_com :
20: em_ov :
21: em_or :
22: em_pat :
23: em_ph :
24: em_pl :
25: em_ro :
26: em_sts :
27: em_sy :
28: em_un :
29: em_vl :
30: em_htgo_hum :
31: em_htgo_inv :
32: em_htgo_rnd :
33: em_htg_hum :
34: em_htg_inv :
35: em_htg_rnd :
36: em_htg_other :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1823.4	99.9	1825	8	SCYR117W	249617 S.cerevisiae
2	1553.4	85.1	1706	8	SCU7137	077137 Saccharomyces
3	1342	18.7	1555	8	SCYR116W	249616 S.cerevisiae
4	195	10.7	99360	8	NCB11N2	AL513444 Neurospora
5	187	10.2	36920	8	SPAC3H1	268144 S. pombe ch
6	137.2	7.5	1275	8	AF353722	AF353722 Arabidops
7	136	7.5	686	6	A86270	A86270 Sequence 92
8	135	7.5	686	6	AR155763	AR155763 Sequence
9	136	7.5	686	6	E66288	E66288 genome DNA
10	121	6.6	935	8	SCYR118C	249618 S.cerevisiae
11	117.4	6.4	1859	9	AF064867	AF064867 Homo sapi
12	117.4	6.4	2963	9	AB016068	AB016068 Homo sapi
13	117.4	6.4	2966	9	HSY13834	Y13834 Homo sapien
14	117.4	6.4	2966	6	E32056	E32056 Human AF01
15	114.2	6.3	1340	3	AF358443	AF358443 Physarum
16	98.4	5.4	78592	3	AC004335	AC004335 Drosophila
17	98.4	5.4	102413	2	AC020024	AC020024 Drosophila
18	92.8	5.1	41337	3	CBRG47019	AC084676 Caenorhab
19	91.6	5.0	40767	3	CEC04F12	281461 Caenorhabdi
20	98	4.8	317511	1	CJ11168X3	AL138076 Campyloba
21	86.6	4.7	582	8	CNS018P	AL110644 Botrytis
22	81.6	4.5	660	8	CNS01AH3	AL112927 Botrytis
23	72.2	4.0	2161	9	AK027874	AK027874 Homo sapi
24	66.8	3.7	1072	6	AX055334	AX055334 Sequence
25	66.6	3.6	124808	2	AP000652	AP000652 Homo sapi
26	65.6	3.6	16488	1	AE004992	AE004992 Halobacter
27	65	3.6	256172	2	AC005139	AC005139 Plasmodiu
28	64.4	3.5	169423	2	AC069309	AC069309 Mus muscu
29	64.4	3.5	186306	2	AC083946	AC083946 Mus muscu
30	63.6	3.5	1245	6	AX055202	AX055202 Sequence
31	63.6	3.5	19715	1	AE000555	AE000555 Helicobac
32	63.6	3.5	234112	3	PFMAL4P2	AL035475 Plasmodiu
33	62.4	3.4	13383	3	AE001374	AE001374 Plasmodiu
34	61.8	3.4	98734	2	PFMAL1P2	AL031745 Plasmodiu
35	61.6	3.4	14943	1	AE001528	AE001528 Helicobac
36	59.8	3.3	12029	3	AE001392	AE001392 Plasmodiu
37	59.8	3.3	110000	2	AC087899	AC087899 Mus muscu
38	59.2	3.2	114736	3	PFMAL3P3	298547 Plasmodiu
39	59.2	3.2	253305	3	PFMAL3P7	AL034559 Plasmodiu
40	58.8	3.2	98899	2	AL603743	AL603743 Danio rer
41	58.8	3.2	318221	2	PFMAL13P3	AL049184 Plasmodiu
42	58.8	3.2	318503	2	AC074166	AC074166 Mus muscu
43	58.4	3.2	172239	2	AC069597	AC069597 Homo sapi
44	58.2	3.2	12029	3	AE001412	AE001412 Plasmodiu
45	58.2	3.2	187469	9	AL357312	AL357312 Human DNA

ALIGNMENTS

RESULT 1
SCYR117W
LOCUS SCYR117W 1825 bp DNA 11-NOV-1997
DEFINITION S.cerevisiae chromosome X reading frame ORF YCR117W.
ACCESSION Z49617.1 J13136
VERSION Z49617.1 GI:1015836
KEYWORDS
SOURCE
ORGANISM

REFERENCE

1 (bases 1 to 1825)
Rose, M., Koetter, P. and Eutlan, K.D.

Unpublished
2 (bases 1 to 1825)

Direct Submission
Submitted (25-SEP-1995) Data collected by MIPS on behalf of the

European yeast chromosome X sequencing project. MIPS at the

Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

601 ttcatcaatacgcctctccctaatactgacacatggccgttcttattgaatgca 660

[illegible]

OY	331	aggccttatctcatgtgtgatctcttaagaagatctctgcacatctctataatccogtggaa	330
Db	61	AGGCGTTTATCTACGTGTTGATCTTAAGAGATCTTCGACCATCTCAATATCCCTGGAAA	120
OY	391	ttaacatcttcgggtctcgcgatgcgccaatctcttcgcagatcttaactgaactagaca	450
Db	121	TTAATCATTTCTGCGTTCGATGGCCAAATTTCTTCGAATCTTACTTGACGTACAGA	180
OY	451	castracagaaagctactbaacaaagatgycacactgycctggaagaagaatgtagat	510
Db	181	CAGTACAGAAAGTATCTGAACAAGTTGCCACTGTGCTGGAAGAGAATTGATGAT	240
OY	511	gaacctttcaataacaaagaaactactccggcgcaaggccaagttccattctcgt	570
Db	241	GAACCTTTTCATAAACAAGAACTACTCCGGGCGAAGGCCAAGTTCCATTTTGGGT	300
OY	571	gagctcatctacttaagcccaaaagcagtttctacaaatcagacactctcccaaatc	630
Db	301	GAGCTCATTAACCTACCCAAAAAGCTAGTTTCATCAATAAGACACTTCCCTPAATC	360
OY	631	tggacataggccgtttctctatgtaatgtagcctgcgcagcaagaattcatatgctcc	690
Db	361	TGGACATAGGCCCTTCTTATTTGAATGCAAGCTCGCCGACATTCATATGCTGCC	420
OY	691	actgctgcacagaaagttaatgctctctggctctctatccagttgctcacttggtgat	750
Db	421	ACTGTCCACAGAGTTTATGCTCTTGGGCTCTTATCCAGTTTGTCACCTTGAT	480
OY	751	tttgcaactctctactatgacattttgcccggagaagaaatttggtttcaataaatg	810
Db	481	TTTGCACTCTCTTACATATACCCATTTTGTCCGGAAGAAAATTTGGTTCAATAATG	540
OY	811	accgtcaactatgatacgaacgaatgatcaagadctcgaacttggtgatgcattggt	870
Db	541	ACGCTCAACTATGATGATCACCGATATGATCAAGAAGCTGCACTTGGCTATATGGT	600
OY	871	ggcccaactccttaactcgtctccttaagatccttgaataatccctacgaattcccttg	930
Db	601	GGCCCACTCTTAACGTCGTCTTAAAGATCTTTGAATAATCCCTAGATTTCCCTTGG	660
OY	931	taactatgctctcttgctgtgtgccaatcttaagcattgaacaataatccagcttc	990
Db	661	TACTATTTAGTCCTTGTGTCGTTGTCACAACTTAAGCATACATATTCACAGTCTC	720
OY	991	atcatgcacatgtttaaagatcactccatctggagagcagtggaactgaaaaatcctc	1050
Db	721	ATCATGCCCATGTTTAAATTAAGTTCACTCCATTGGAGAGACGTGAACCTAATAATCTATT	780
OY	1051	gaagatttggccgatalaagatgtggtctccctctagaataagatttgtloattgacggtca	1110
Db	781	GAAAGTTTGGCCGATAGAGTTGGGCTTCCCTAGATTAAGATTTTGTCAATGAGGGCTCA	840
OY	1111	aaaagatctctcaatccaacgcgcatcttccaagtttgccattcaactccaagagaatt	1170
Db	841	AAAAGATCTTCTCATTTCAACGCAATTTTCAACAGGTTTCCCTTCCCAAGAGATT	900
OY	1171	gtttgtctgcacacttaaggaacgagtaattctcaatgaaatgaagcgttttggcc	1230
Db	901	GTTTGTCTGCACACTTATAGTGAACATATTTCTACTGATGAATTAACGGCTGTTTGGCC	960
OY	1231	catgaaatcggtaactgycgaaaaaaacacatcgttaatalatgtaactcttaagtaactg	1290
Db	961	CATGAATTCGGTCACTGCGCAAAAAACCAACATCGTTAATAATGTCATCTTAAGTCAATTG	1020
OY	1291	caaaccttctctaatcttctccctttccacagcatctcaagaataaactcatctttaaac	1350
Db	1021	CACACCTTCCCATTTTCTCCCTTTTACACACACTCTACAGAAATACATATTTCACAC	1080
OY	1351	accttgagctttctcttaagaaagctcactgagcttggttgatcccgattacaacag	1410
Db	1081	ACCTTGCGCTTTTCTTAAGAAAGTCACTGGCAGTTTGTGTGATCCCGTTATGCACTRAG	1140
OY	1411	gaattcccatatcatatgattatcatatgatttaacgacttaataaccacactcgaatgt	1470

THIS PAGE BLANK (1.3PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460A-3

Perfect score: 2948

Sequence: 1 tgaactcttctgatacaaaag.....gggagagataaagaatcaca 2948

Scoring table: IDENTITY-NUC

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 45 summaries

GenEmbl:*

- 1: gb.ba:*
- 2: gb.hg:*
- 3: gb.in:*
- 4: gb.om:*
- 5: gb.ov:*
- 6: gb.pat:*
- 7: gb.ph:*
- 8: gb.pl:*
- 9: gb.pr:*
- 10: gb.ro:*
- 11: gb.sts:*
- 12: gb.sy:*
- 13: gb.un:*
- 14: gb.vl:*
- 15: em.ba:*
- 16: em.fun:*
- 17: em.hum:*
- 18: em.in:*
- 19: em.om:*
- 20: em.or:*
- 21: em.ov:*
- 22: em.pat:*
- 23: em.ph:*
- 24: em.pl:*
- 25: em.ro:*
- 26: em.sts:*
- 27: em.sy:*
- 28: em.un:*
- 29: em.vl:*
- 30: em.hggo.hum:*
- 31: em.hggo.inv:*
- 32: em.hggo.rod:*
- 33: em.hggo.hum:*
- 34: em.hggo.inv:*
- 35: em.hggo.rod:*
- 36: em.hggo.other:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	2948	100.0	29987	8	SC8156	249260 S.cerevisia
C 2	1560.4	52.9	4156	8	D50276	50276 Saccharomyc
C 3	1361	46.2	4130	8	YSCNR1G	142821 Saccharomyc
C 4	1163	39.5	4225	8	SCU65682	U65682 Saccharomyc
C 5	1069.4	36.3	3250	8	SCU32580	U32580 Saccharomyc
C 6	931	31.6	3525	8	SCU63849	U63849 Saccharomyc
C 7	786.8	26.7	3240	8	YSCBU1	D50843 yeast BULL
C 8	703.8	23.9	3020	8	SCRD1G8N	X88901 S.cerevisia
C 9	225.2	7.6	270	8	YSCGNC8X	N87379 yeast Eco R
C 10	143.8	4.9	27559	8	SC8339	249210 S.cerevisia
C 11	75	2.5	2023	8	SCYR066C	271683 S.cerevisia
C 12	71.8	2.4	4775	8	SCU32938	U32938 Saccharomyc
C 13	68	2.3	780	11	CNS06K8B	AL402561 T7 end of
C 14	65.6	2.2	256172	2	AC005139	AC005139 Plasmodu
C 15	65.6	2.2	310779	2	AC005140	AC005140 Plasmodu
C 16	64.4	2.2	904	11	CNS06189	AL399967 T3 end of
C 17	64.4	2.2	86827	3	PFMAL3P5	AL034556 Plasmodu
C 18	63.4	2.2	1920	3	PFMAL3P5	X62393 P.falciparu
C 19	63.4	2.2	110000	2	PFMAL13P2_0	AL034559 Plasmodu
C 20	62	2.1	253305	3	PFMAL13P7	AL034559 Plasmodu
C 21	61.6	2.1	1141	6	AX083744	AX083744 Sequence
C 22	60.8	2.1	234112	3	PFMAL4P2	AL035475 Plasmodu
C 23	60.6	2.1	202872	2	AC016160	AC016160 Homo sapi
C 24	60	2.0	5632	3	AF310892	AF310892 Dictyoste
C 25	59.4	2.0	3214	3	AF019980	AF019980 Dictyoste
C 26	59.4	2.0	318221	2	PFMAL13P3	AL049184 Plasmodu
C 27	58.8	2.0	321003	2	PFMAL4P3	AL035476 Plasmodu
C 28	58.4	2.0	78439	17	HS085197	U85197 Homo sapien
C 29	58.4	2.0	169546	2	AC004157	AC004157 Plasmodu
C 30	58.4	2.0	245802	2	AC006279	AC006279 Plasmodu
C 31	58.4	2.0	251124	2	HD000660	HD000660 Homo sapi
C 32	58.2	2.0	153098	3	PFMAL3P2	AL034558 Plasmodu
C 33	58	2.0	159724	2	AC004688	AL034558 Plasmodu
C 34	58	2.0	178157	9	AC011302	AC011302 Homo sapi
C 35	57.8	2.0	13684	3	AE001403	AE001403 Plasmodu
C 36	57.8	2.0	153477	2	AC006278	AC006278 Plasmodu
C 37	57.6	2.0	53932	2	AC023371	AC023371 Homo sapi
C 38	57	1.9	12029	3	AE001431	AE001431 Plasmodu
C 39	57	1.9	153443	2	AC006280	AC006280 Plasmodu
C 40	57	1.9	178273	2	AC005308	AC005308 Plasmodu
C 41	57	1.9	196149	2	AC004709	AC004709 Plasmodu
C 42	56.8	1.9	67970	3	PFMAL1P2	AL031746 Plasmodu
C 43	56.8	1.9	149627	9	AC087428	AC087428 Homo sapi
C 44	56.6	1.9	92633	2	PFMAL4P1_3	Continuation (4 of
C 45	56.6	1.9	153477	2	AC006278	AC006278 Plasmodu

ALIGNMENTS

RESULT 1

LOCUS: SC8156/c

DEFINITION: S.cerevisiae chromosome XIII cosmid 8156.

ACCESSION: 249260.1

VERSION: 249260.1

KEYWORDS: BULL1, cyclochrome b5, DAG1, Initiation factor 1A, Inorganic pyrophosphatase, Ipp2, nitrate reductase, Orotate phosphoribosyltransferase, tau element, TIF1, PRS3, transfer RNA-Gln, trehalose-phosphate synthase, U6 snRNP, URA10, baker's yeast.

SOURCE: Saccharomyces cerevisiae

ORGANISM: Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 29987)

AUTHORS: Lye, G. and Church, C.M.

JOURNAL: Unpublished

REFERENCE 2 (bases 1 to 29987)

AUTHORS: Barrett, B. and Rajandream, M.A.

TITLE: Direct Submission

11-AUG-1997

JOURNAL

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CA10 1RQ E-mail: barrell@sanger.ac.uk

COMMENT

All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis.

Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)
is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry
SC9920,
accession no. Z48639 and at the end by cosmid 8021.

FEATURES

source

1. 29987
Location/Qualifiers

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 8156"

/map="13R"

/complement(<1..356)

/note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps
and extends YM9920.13c"

/codon_start=1

/product="unknown"

/protein_id="CA89242.1"

/db_xref="GI:809082"

/db_xref="SWISS-PROT:O03496"

/translation="MTDVFSSQSEIEFRVKEFLIANNPAKINNENKDAVTOIEHD
PRVLYQKDDLPNLESTRLIFPDFTICLLRSQIILASKKIDQEFLSAKRELLT
EANNITFEYVLDFA"

1..140

/note="overlap with SC9920 Z48639 S.cerevisiae chromosome
XIII cosmid 9920"

/complement(851..1312)

/gene="TIF11"

/complement(851..1312)

/gene="TIF11"

/note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891,
initiation factor 1a"

/codon_start=1

/product="TIF1P"

/protein_id="CA89243.1"

/db_xref="GI:809083"

/db_xref="SWISS-PROT:P38912"

/translation="MGKNTKGGKRGKRGKNDGPKRELIYKEGGOEYAOITKMLGN
GVEASCPDGNKRMHRTGKLRKTYMGGDIIIVSLADPDODDGDVYHKYLDART
LKNQELPENAKINETDNFGFESDDVNFEPGNADDEDEGEDELIDDT"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST P38426, alpha.alpha-trehalose-phosphate
synthase and PIR:S27471; PS00213 lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"

/codon_start=1

/product="TPS3P"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIYASLFLPTQFEADVNTSDAKIVSSMTIKVDDNNOELS
NNKRRSSSVTSASHYIGLPEAQINGEPLORANTGSPATGVYTHNEMHSSQFL
ELTANATHAASGILPANNPVSSGTAQRPVSEEFSAVSARVCSPEASASSISA
SRSSAHNDLSSLMKNPNTSFDSPRPVRSKSSKVIIPVSKSVDDPAVADVAKY
REEFOQASLSPMKRVSGSTAGDSIASSSNRLYSQGFQDNFIEDTDEDDIDSLD
TDATKYVPRKGGYNNAKLRLMRNRYELFKHLPTIYDSGNGSLKRAVNIAY
AKTYKEPVSWGTMGIPRDELPHVCHIKSKLEDDSSFPVYDDITFKAYKNYA

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

misc.feature

CDS

/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously
noted ORF, SW:YTP3_YEAST P38430"

/codon_start=1

/product="unknown"

/protein_id="CA89245.1"

/db_xref="GI:809085"

/db_xref="SWISS-PROT:P38430"

/translation="MNLVDACHVITDPONTFCGDDGSGGTLRCVSSNPTDNNLT
KLAGRSTSKNDICVGPVHPVSHLYVGSRRDVSHTQVLYKKEGSDSLVQVL
PEPLDEEYIKREFNDLVAVIGIGLCKFLRLANVFYMNENKALITTVKTLKSHOE
TVPRRFRRLARHTSKPISIDVQCHGDLINCEBELLYHSYKICLHSYTSKELLG
QWLKPPRDLFVLSKWINFKDPEBDALVRSYPCILLETVDYPIIDNPDSYKAL
TEQIYINAOIARADMTLDAQAALRYENFQCFIK"

6392..6997

/note="YM8156.05, unknown, len: 301, CAI: 0.12,
SW:YTP2_YEAST P38429 previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89246.1"

/db_xref="GI:809086"

/db_xref="SWISS-PROT:P38429"

/translation="MARPVYNAETESRGRTGCGVYASNNNGSCNNNGSNNNNNN
NNNNNSNNNNNGPTSSGRTNGKOLAQOQYIKRLIETHIDNPDLAPKSHPM
DPEVYDAPLRYVDKHDOLQVDPDLTQGYLGSIKAKYTSYKRNNGQGHKRLHKK
DLANVVRHDEHSIKETDCIPQFTIYVKNKKFKKEFRG"

7278..7889

/note="YM8156.06, unknown, len: 301, CAI: 0.13,
SW:YTP1_YEAST P38428, extends previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89247.1"

/db_xref="GI:809087"

/db_xref="SWISS-PROT:P38428"

/translation="MEDSRLLITLILVGVFLKFFQSNQHPASAQRTSATGVNAHR
PGSTQNALPRTGVNGHPVPTQVYVQMLAPLHEDQRTSYSELTNGSVSEYVRY
LRGEFSPPEPGEPSRAPGCMANAVDNNAGGGEENDPRKMKIACENLIDFHYDLN
EDMSNLSEFKDIDIERKRLVWQAARKLETQLQSDKQSLST"

/complement(8013..9398)

/note="YM8156.07c, unknown, len: 461, CAI: 0.11"

/codon_start=1

/product="unknown"

/protein_id="CA89248.1"

/db_xref="GI:809088"

/db_xref="SWISS-PROT:O03508"

/translation="MEPEEFRRRGGSSRCGNHRYLRKNSACELAVPEVPEAI
TKLIYDQDNLTFNTPGTEDLYRELNLITSLTPNGGNNNEGFLQAAIEISTEK
PRRSWMDADIVKLAESTISADOTISIVTGRSEEFKHLIEHALDTANSHMCSENEF
RFNAVCLAKRAISEYTSKRYKRELMDELEVYPSLSELSYDRIHQIDAFKSFSLD
LPRLKWSAIPVPRPTKSLPRLQELMDVDMYRKNSQALSTQKDLRPTPQIGTIL
CTAHRSLISIVIKRYLKRGRGRTFRPETHYHPYIPCAEFGKDPALAEIAYVMS
TRPDSEKNOVHISOITYLPDGCIVHFOYTDLAVISAHHNRKPELEVYKAPADP
NRTFTLEPEIYVGHYTKRRIEDLEVYTERLLNCKEDIMHVPDNTIPIAIFGRF
AKLAIPKNSA"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"

/codon_start=1

CDS

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"

CDS

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"

CDS

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"

CDS

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"

Query Match	100.0%	Score 2948	DB 8	Length 29987
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2948	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	tgaagctttgaatgacaaagaagaagctcgacaagcatcaagaagctctgtaagatattcca	60
Db	27785	TGAAGCTGTGATGAAACAAGAGAAAGCTGCACAGATCTAAAGCTTGTGTCGATGATTTCCA	27726
QY	61	ttcaaaaattctgtaatacgaaaccaagtctcaacaagaattctctgaattaataatgaatt	120
Db	27725	TTCAAAAATTGTGTAATATGAAACCAAGTTCAACAAGATTTTCTTGAAATTAATGAGTT	27666
QY	121	atataatgaaatgggggaagacgtaaggccaagaagaaactgaattatcaagattatlaac	180
Db	27665	ATAAATAATGAAATGAGGGAGACCGTAGGCCAAAGAACTGAAATTTACAGATTTATATAC	27606
QY	181	ttcaacagctggttaacgatatcgaaagaagcatttgcaacttgaagttaagtgttcaacatt	240
Db	27605	TTCAAGCTGTTTACGATATCGAAGAGATTTGCAACTGAAAGTTAGTGTTCACAACTT	27546
QY	241	atccaaacatttttaaaaaaacaggttcagtaaccttaaaaacacactccaagaacgatatc	300
Db	27545	ATCCAAATTTTTTAAAAAACGGTCAGTACCTTTAAACAAACATCAAGACGCAATTATC	27486
QY	301	tgagatcatcaatcatcgacacacaggtlaagggtaagtcacatcgctcgccacagttcaagctaat	360
Db	27485	TGAAGATTCAATATCGCACACACAGGTAAAGGTATGATCATCGCGCCAGTTAAGCGTCAATT	27426
QY	361	aagcgccagtaaactcttcacccaagaagtagttatttttttaacctagcggtagctcgtctac	420
Db	27425	AAGCCACGTAATCTTCTTCATCCAAAGAGATGTTTATTTTATACCTAGCGGTACTCGTCTAC	27366
QY	421	ttccgctaaatttcacagccagatattgttcataaaatggttgagattgctcctttaacgta	480
Db	27365	TTCCCTGAATTTACAGACCGAGATTGTTCATTAATAGGGTTAGATGCTGCTCTTACAGATA	27306

QY	481	caacagagacattaaatgtgaacttggaaatttaataagaacattaaggaaaccttaattcc	540
D	27305	CAAAACGAGCATTAATGAATGTGAACCTGGAAATTAAATAAGACATTAAAGCAAACCTTAAATTC	27246
QY	541	aagctttgaaagcgcctcatgttgtaggttttatltgcttcgagtaaagttaattga	600
D	27245	AAGTTTGAAAGCTGCCTTAGTTGTGAAGTTTTATGTGGTTCCGATAATGATTAATTTGA	27186
QY	601	aaaccatcttggtgtagcagaagaattgatcccatcttccgttaggcgaagtacaanaa	660
D	27185	AAACCATCTTGGCGTAGCAGAAGATGTGATATCCCATTTCTGTATGGCAAGACGACAATAA	27126
QY	661	aaaaacattagaaaaaatctcgtaactttctctatagatatagatatgtatgtt	720
D	27125	AAAAAACATTAGAAAAAAAATCTCGTTACTTTCTTATAGATATAGATATATGTATGTGTTT	27066
QY	721	gcttatagatgaaggatattatcgacgccctttatcccattatataaanaattctt	780
D	27065	GCTTATAGATGAAGATTAATTAATGCGCGCTTGATATCCATTAATTAATAAATTTCTTT	27006
QY	781	taaaatgcatttctggtgtccttttgttgccttcgtatatttctttcttgagcaactg	840
D	27005	TAAATGCATTTCTGTGTGCTCTTTGTGTCTGTCTGTATTTTTTTTTTTTGGACCACTG	26946
QY	841	gatygnaaaccttgbatgatltctaacttaactttaagtlaactaaataatogagatt	900
D	26945	GATGGAAGAAACCTTGAATGATTTTAATCACTTTATTTAACTTAATAAATATCGAGATTT	26886
QY	901	caggaacaaacacatagatttcttcttgtagaanaaaataaagaanaataattgatgtc	960
D	26885	CAGGAACAAACATAGATTTCTTCTGTCAAGAAAAATAAAGCAATAAATATGATCTT	26826
QY	961	tgactactgacttctgtcataagagaacacagaacgaactgtacaatttccaacatt	1020
D	26825	TGACTACTGACTCTGTCTCATATAGAGAAACACAGAACGCAATGCTCAATTCACAAATT	26766
QY	1021	tctagtgctccatacatctccatctcatalgtgtacgcgtalatgycaaacttacaacc	1080
D	26765	TCTAGTCTCTCTATACATCTCCATATCCATATCTATGTGCTCCGCTATATGCCAATTCACAAACC	26706
QY	1081	agaaaggtctcaaacagagataractctcgaaagatattaatctcgatgtaaaaaacttaaat	1140
D	26705	AGAAAGGTCTCAAAAGAGATATCTCTCGAAAGATTAATCTCGATGTGAAAAACCTTAACAAT	26646
QY	1141	tatgctaatttccaacaccttcttctgtgctcttcttcaatctccaattactcagtacac	1200
D	26645	TATGCTAATTTCCAACCTTTTTTGGTGCTTTTTTACATATCATAATATCTAGTNACCAC	26586
QY	1201	tccaacataaagtttcaaaggacatlttcaaggcttagtataatcacccaggttaatacgc	1260
D	26585	TTCCACATATATAGTTTCAAGAGCACCAATTTTAGGCTTAGGATATATATCCACAGGTATATTACGC	26526
QY	1261	tgcattgcaccaaacctcttggaatctcaagccaggttcgtgaagaagtcttaacgaatgtgtgc	1320
D	26525	TGCATTGCCAAACCTCTTGCAATTCAGCCAGTTCTGTAAAGAACTTACGAAATGTGTGTC	26466
QY	1321	gabtattatgacctlatatgttgyaacccgttttagaatttggitatitatalcaattatnaa	1380
D	26465	GABTATTATGACCTTATATGTGTGACCCCGTTTAGATTTTGTATATATCATTTATTTAAA	26406
QY	1381	tccaagaagcctcataactgtgaagaatttttaaccatgaaatccctgaaataatttgaagcttcag	1440
D	26405	TCCAAGAAGCTCATCTTGGAAMATTTTTCACATGTAGATTTCTCAAAATATTGGAGTTTCAG	26346
QY	1441	gaatttatatttggacaacaaataactaggaataatttttaacgttcaatgtctttgactac	1500
D	26345	GAAITTTATATTTCACACCAATFACCTGAGGAAATATTTTACACGTCAATGCTTTTGACTTAC	26286
QY	1501	gtacttaaacctlaataccgcatctgcgaactaagctatcdaacagttlaatttgcgaacacat	1560
D	26285	GTACTTAAACCTAATATCCGCAATTCGCAACATTAACCTATCAACAGATTTATTTTGGCAACATTC	26226
QY	1561	gccttttttttgagactgagcagcagcaaccaatgcttatagcaattacaggaagctccat	1620

Db 26225 GCTTTTGGACCTTGGCCGACACACACCTTATAGCAATTAAGGAAAGGTCAT 26166
 Qy 1621 gacacgtttccatctctgctgacacacatgcttccaaattatcacacacacttttgg 1680
 Db 26165 GACACACGTTCCATTCGCGACACACATGCTCCAAATTTTATACACAAACCTTTTGG 26106
 Qy 1681 aggttaacacagatttgcattcgttaagaacagcgaggaaactaagtgctgcataactc 1740
 Db 26105 AGGCTTAACCAAGTTGATTCGTAAAGACGCGGAAACCTATGCTCGATTAATCCT 26046
 Qy 1741 gcaagcccttggcaataatcagggggttccctggtcctcaagattgaattacattac 1800
 Db 26045 GCAAGCCCTTGGCAATATCATAGGGGTTCCTGGCTTCACAGATTAATTAATTAATTCAC 25986
 Qy 1801 agatgaacacagaagaagctgagacgacattcccaattgtctcacaactggaataagtgta 1860
 Db 25985 AGTATGACACAAAGAAAGCTGGACGATTCGCAAAATGGCTCAATCGGAATTAATGCTA 25926
 Qy 1861 ctgcgactgctgctccttgatataatccctgaagagatcactacaactcgtgtagg 1920
 Db 25925 CTGCGACATGCTGCTGCTGATTAATCCCTGAAGGATTAACCTTAACAAACTCTGTAGG 25866
 Qy 1921 aactcctggtatagaataaccccttagcctttttagtactgtatagcgtttaaact 1980
 Db 25865 AACTCCTGTTATAGAAATTAACCTTAGCCTTTTACGTAAGTAACCTTAATTAAT 25806
 Qy 1981 ttccatgctacataacaccttttccactactatgaaactcagacgacgagct 2040
 Db 25805 TTCCATGATACATTAACCTTTTTCACATCTATTAAGAAATCTAAGACGACGCGCT 25746
 Qy 2041 ttgttgaagagagtgaaaaacacagcttctggtcttggctgaaagaatttggagac 2100
 Db 25745 TTGTGACGAGAGATGAAGAAATGACAGTTTGGTGTGGGAAAGAAATTTGGAGGAC 25686
 Qy 2101 tataagaacacataacttctgattagagactcaataacagctcgtctgaagtgat 2160
 Db 25685 TATAAGTACCTTAATCTTGTATACGAGACTCAATTAACAGCTGCTGCTGAT 25626
 Qy 2161 tgaagctgacagatcctaaagagagagagagagagagagagagagagagagagagag 2220
 Db 25625 TGAAGCTGACATCTAAG 25566
 Qy 2221 tttcttgaagcttcttctgctcctgaatctaaagctcctgaagcttctgcttct 2280
 Db 25565 TTTTCTTAAGCTTCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25506
 Qy 2281 ttgtggtgtacatatttcaattcaagaagagagagagagagagagagagagagagag 2340
 Db 25505 TTGTGCTGTACATTTTCAATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25446
 Qy 2341 agagataacag 2400
 Db 25445 AGAGATTAACGAG 25386
 Qy 2401 gataaacggaagctcgaagcttctgattgctgcacagctccctgcaaatgaaatccgacg 2460
 Db 25385 GATTAACGGAAGCTGAAGTTTGAATGCTGACAGAGCTGCTGACAAATGAATCCGACG 25326
 Qy 2461 gtaaaaaaacctaaag 2520
 Db 25325 GTAATAAACCTTAAG 25266
 Qy 2521 gataataattggttgag 2580
 Db 25265 GATTAATAAATCGGTGGGAG 25206
 Qy 2581 tctgcgtcaatgcaagctgac 2640
 Db 25205 TCTGCCTCATGCGAG 25146
 Qy 2641 ccgtatagaagacttgcagtgag 2700

Db 25145 CCGCTAGAGAACTTCATGGAGGAGGTAACACAGAGATAGAAATCTCCAAATAGACTAAA 25086
 Qy 2701 caaggtactactaggtataaaaaaggggtgtacactccatccaggaattaaatgct 2760
 Db 25085 CAAAGTAACTAATGATTAATAAAAAAGGTTTCACTCCATCCAGGAAATTAATGCT 25026
 Qy 2761 aagctataaagaacacttatactgttccgcacatcacacacacacacacacacacacacac 2820
 Db 25025 AAGCTAATTAAGAAACTTAATTTATGGGTTCCCGCAATCAACACCTTAAGCTTAAGCT 24966
 Qy 2821 gataattccatagagcttgcacagatcattacaataatatacaactaagcagacatgct 2880
 Db 24965 GATAATTTCCATAGAGCTGTACAGATTAATCTTAACAAATATACAACTAAGCAGCAATGCT 24906
 Qy 2881 gaagataatgctggaatagacatgaataacacacacacacacacacacacacacacacacac 2940
 Db 24905 GAAGATATATGATGGAGATTAAGCAATTAATTAAGATTAAGGATTAAGGAGATTA 24846
 Qy 2941 gaatcaca 2948
 Db 24845 GAATCACA 24838

RESULT 2

LOCUS D50276 4156 bp DNA PLN 10-FEB-1999
 DEFINITION Saccharomyces cerevisiae HST1 gene, complete cds.
 ACCESSION D50276
 VERSION D50276.1 GI:1777313
 KEYWORDS HST1; high-copy suppressor of TFP sensitivity.
 SOURCE Saccharomyces cerevisiae (strain: RAY-3A) DNA.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.

REFERENCE

1 (bases 1 to 4156)
 Tsuchiya, E.
 Direct Submission
 Submitted (13-APR-1995) to the DDBJ/EMBL/GenBank databases. Eiko
 Tsuchiya, Hiroshima University Faculty of Engineering, Department
 of Fermentation Technology, Kagami-yama, Higashi-Hiroshima,
 Hiroshima 739, Japan (E-mail: etsuchi@ipc.hiroshima-u.ac.jp,
 Tel: 0824-24-7868, Fax: 0824-24-7868).

REFERENCE

2 (bases 1 to 4156)
 Tsuchiya, E., Matsuzaki, G., Tsukao, A. and Miyakawa, T.
 Molecular cloning and characterization of S. cerevisiae HST1 gene
 3 (sites)
 Unpublished (1995)

AUTHORS

TITLE

The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance
 to high concentration of Ca²⁺ with the participation of
 HST1/NAC1/BFR1
 Gene 176 (1-2), 35-38 (1996)

JOURNAL

MEDLINE

FEATURES

source

Location/Qualifiers
 1..4156
 /organism="Saccharomyces cerevisiae"
 /strain="RAY-3A"
 /db_xref="taxon:4932"
 946..3693
 /gene="HST1"
 946..3693
 /gene="HST1"
 /standard_name="high-copy suppressor of TFP sensitivity"
 /codon_start=1
 /protein_id="BAA0819.1"
 /db_xref="GI:1816450"
 /translation="WSNRDNESMLFTSSDKAIAQRDKRSEVLIQAASLNEIRSY
 KNLKRLSGSDLLIDPEIDIFGESSGRSWSGTTSSSAMPDSTTVNTRISDP
 TPLENLRGNSGDISSNKTGAGYGLIKGVHSPSRKLANVLLKMLVAVANQHP
 VKPNFLVQDTLQIOLSDNGEDNGNSNNDIEDNGEKSQSYENKENTINTL
 NRGSLRHGNASLIRPSTLRBSYTFEDNEDNDNDGDSASETVNVEERISIKERP
 SUDRTIEELTKISNAGLNDNDATILARLSNAGSYSDKKDQPOPEGHIDEDDIDGST

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(Without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460a-3
Sequence: 1 tgaactgtatgaacaagaag.....ggggaggaataaatacaca 2948

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

GenBank: 1: gb_ba: 2: gb_hgt: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_om: 20: em_or: 21: em_ov: 22: em_pat: 23: em_ph: 24: em_pl: 25: em_ro: 26: em_sts: 27: em_sy: 28: em_un: 29: em_vl: 30: em_hgt_hum: 31: em_hgt_in: 32: em_hgt_rod: 33: em_hgt_hum: 34: em_hgt_in: 35: em_hgt_rod: 36: em_hgt_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2948	100.0	29987	8	SC8156
2	1560.4	52.9	4156	8	D50276
3	1361	46.2	4130	8	YSCNCRIG
4	1163	39.5	4225	8	SCU65682
5	1063.4	36.3	3250	8	SCU32580
6	931	31.6	3525	8	SCU63849
7	786.8	26.7	3240	8	YSCUBJ1
8	703.8	23.9	3020	8	SCRDSTGEN
9	725.2	7.6	270	8	YSCGWC58X
10	143.8	4.9	27559	8	SC8338
11	75	2.5	2023	8	SCYNR06C
12	71.8	2.4	4775	8	SCU32938
13	68	2.3	780	11	CNS06K8B
14	65.6	2.2	256172	2	AC005139
15	65.6	2.2	310779	2	AC005140
16	64.4	2.2	904	11	CNS06T89
17	64.4	2.2	86827	3	PFMAL3P5
18	63.4	2.2	1920	3	PFMAL3P7
19	63.4	2.2	110000	2	PFMAL3P2_0
20	62	2.1	253305	3	PFMAL3P7
21	61.6	2.1	1141	6	AX083744
22	60.8	2.1	234112	3	PFMAL4P2
23	60.6	2.1	202872	2	AC016160
24	60	2.0	5632	3	AF310892
25	59.4	2.0	3214	3	AF019980
26	59.4	2.0	318221	2	PFMAL3P3
27	58.8	2.0	321003	2	PFMAL4P3
28	58.4	2.0	78439	17	HS085197
29	58.4	2.0	169546	2	AC004157
30	58.4	2.0	245802	2	AC006279
31	58.4	2.0	251124	9	HUAE000660
32	58.2	2.0	153098	3	PFMAL3P2
33	58	2.0	163794	2	AC004688
34	57.8	2.0	178137	9	AC001302
35	57.8	2.0	13684	3	AE001403
36	57.8	2.0	153477	2	AC006278
37	57.6	2.0	53932	2	AC023371
38	57	1.9	13029	3	AE001431
39	57	1.9	163443	2	AC006280
40	57	1.9	178273	2	AC005308
41	57	1.9	196149	2	AC004709
42	56.8	1.9	67970	3	PFMAL3P3
43	56.8	1.9	149627	9	AC087428
44	56.6	1.9	92633	2	PFMAL4P1_3
45	56.6	1.9	153477	2	AC006278

ALIGNMENTS

RESULT 1
LOCUS SC8156/c
DEFINITION S.cerevisiae chromosome XIII cosmid 8156.
ACCESSION Z49260.1
VERSION Z49260.1 GI:809081
KEYWORDS BULL1; cytochrome b5; DNGL; initiation factor 1A; inorganic pyrophosphatase; IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF1; TP33; transfer RNA-Glu; trehalose-phosphate synthase; U6 snRNP; URA10.
SOURCE
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 29987)
AUTHORS Lye, G. and Churcher, C.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29987)
AUTHORS Barrett, B. and Rajandream, M.A.
TITLE Direct Submission

THIS PAGE BLANK (OPTO)

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ. E-mail: barrell@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry SC9920, accession no. Z48639 and at the end by cosmid 8021.

FEATURES

source

1. 29987
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8156"
/map="13R"

CDS

complement(<1..356)
/note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YW9920.13c"
/codon_start=1
/product="unknown"
/protein_id="CA89242.1"
/db_xref="GI:809082"
/db_xref="SWISS-PROT:Q03496"
/translation="MTDVQSSQSEIELFRYKEFLIANPKINNENKDAVLQIEHD ERYLIQIKDGLPMLNEDSLRIFPDSICILRSHQIILSKKIDSOEFLSIVKQQLLT EANNIIFEVYLDLFWA"

misc_feature

1. 140
/note="overlap with SC9920 Z48639 S. cerevisiae chromosome XIII cosmid 9920"
complement(851..1312)
/gene="TIF11"
complement(851..1312)
/gene="TIF11"
/note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, Initiation factor 1a"
/codon_start=1
/product="Tif1p"
/protein_id="CA89243.1"
/db_xref="GI:809083"
/db_xref="SWISS-PROT:P38912"
/translation="WGKRTGKGRGKGRGKNDSDGPKRELLYKEGGEYAOITMLGN GRVESCDCGNKRAHRIKGLAKTYMGCQDIIIVSLDPODDCDVYHKYKNDLEART LKNGELPENAKINETDNGFESDENVFEGMDEDEDEGDELIDDI"

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/gene="TPS3"
/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38425, alpha, alpha-trehalose-phosphate synthase and PIR:S27471, P500213 Lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLFLPTTPOFEADVNTSDAKLVESMILKVDCCNOELS NNKORSSSTASSHYIGLPQEOAINGEPLRANVAGSPATGVTHNEMEMLSSEOLF BELTANATHAASGIPRANPYSSGSTRAPRVEVEFFSAPARVCSPOEASASISA SRSASHNDLSSSLMKNPVLSFDSHPVRSSKSVITPVKSVDDPAVDVAKY REFEOQASLPSMKRVSGTADSSIASSSNMLRISQOFQDNFIDTDESDIDLE TDAIKRYNPKGKYSNNAKLRASLRNRYELFKLHPMTIVDSKNGSKNAIVIAV AKTYKEPVSWGTGIGIPDELPHVEVCHRSIKKLEDDFSFPVTDITFKAYKNA

CDS

complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/gene="TPS3"
/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38425, alpha, alpha-trehalose-phosphate synthase and PIR:S27471, P500213 Lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLFLPTTPOFEADVNTSDAKLVESMILKVDCCNOELS NNKORSSSTASSHYIGLPQEOAINGEPLRANVAGSPATGVTHNEMEMLSSEOLF BELTANATHAASGIPRANPYSSGSTRAPRVEVEFFSAPARVCSPOEASASISA SRSASHNDLSSSLMKNPVLSFDSHPVRSSKSVITPVKSVDDPAVDVAKY REFEOQASLPSMKRVSGTADSSIASSSNMLRISQOFQDNFIDTDESDIDLE TDAIKRYNPKGKYSNNAKLRASLRNRYELFKLHPMTIVDSKNGSKNAIVIAV AKTYKEPVSWGTGIGIPDELPHVEVCHRSIKKLEDDFSFPVTDITFKAYKNA

misc_feature

complement(2260..2301)
/gene="TPS3"
/note="P500213 Lipocalin signature"
/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:YTP3_YEAST P38430"

CDS

complement(2260..2301)
/gene="TPS3"
/note="P500213 Lipocalin signature"
/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:YTP3_YEAST P38430"
/codon_start=1
/product="unknown"
/protein_id="CA89245.1"
/db_xref="GI:809085"
/db_xref="SWISS-PROT:P38430"
/translation="NKKLVDAHCHVITDPEDNTEFCGDDGSGQGLRCVMSNPYDMMNL KLAGSTSKNDIGVGFVHPVSHLEYSGSRDKVSHODVLEYENRQFDSLYOVL PPDLEEXIKREPDNLYSVIGIGIDLPFLPANGFOMNERKALITVYKLSHOE TVRRFRWLARHRSKPSISHDYCKHCKLDICNEELLTHSVKICASHSTYKTLIG QWLKRFPPDRIFVSLSKWLNFDPEGDVLVSLPTCLITETDYPIDNPDPSYOKAL TEOLQYLNQOIRAWDETLDASQALRYENRQEFK"

CDS

6392..6997
/note="YM8156.05, unknown, len: 301, CAI: 0.12, SW:YTP2_YEAST P38429 previously noted ORF"
/codon_start=1
/product="unknown"
/protein_id="CA89246.1"
/db_xref="GI:809086"
/db_xref="SWISS-PROT:P38429"
/translation="NAPRYTNAETSSRGRPTGCGGYASNNNSCANNNSNNNNNN NNNNNNSNNNGPTSSGRTNGKOLTAQOOYIKNLITHTIDHDPDLSHPH DPEYTDAPLRKRYKFDVLDQNLGLGILGSKYKRYSTKRNQOGDHRKIRK DLANVVRHREHDSIKETDICIPOFIYKVKKKKFKFEPFG"

CDS

7278..7889
/note="YM8156.06, unknown, len: 301, CAI: 0.13, SW:YTP1_YEAST P38428, extends previously noted ORF"
/codon_start=1
/product="unknown"
/protein_id="CA89247.1"
/db_xref="GI:809087"
/db_xref="SWISS-PROT:P38428"
/translation="MEDSRLLTLLIVFVIFLKFQSNQHPAQRSLATGVNAGR PCOSTONALARTGVNGHPTVOMETVONLAPNLHPEDIRSLSEYGVSELYERT LGDESEFPGEPSRAPGANAADVNNAGGEGPFRKMKCAEYTLDFHVDLN EDMNSLTFDQDIEERRRLIWOARKKLEFKLQSDKQSLLT"

CDS

complement(8013..9398)
/note="YM8156.07c, unknown, len: 461, CAI: 0.11"
/codon_start=1
/product="unknown"
/protein_id="CA89248.1"
/db_xref="GI:809088"
/db_xref="SWISS-PROT:Q03508"
/translation="MEFEDEFRRKGRSSCGNRVLRKNNSCACELAVPEVEBEAI TKLIYDFDNTLFAPTGPTQLTYRELLNLTSLTPNGGWNPGFLOALIEISRTK PRRYSWADIVKLAEESYSKADITISYLGREESKFKLIEHALOTAFSHKSCSENE RPNVAVCAKRAISSEYTSKKKELMBDLEVPYSLRELSYDDRIHOIDAFSEFSEID LPRKNSAIVPRPFTKSLPREQDELAMVWRKNNSCALSTSGQFDLARTPRQICITL CTASHRLSLEVIKYLKRRGRGRTFRKLIEHPYLICAEAGKQIPALELTKVSNND TRPDSSEKQVHISQIYLDPQKCIYHFOVTVDVVASAHNNRKLPLEVFKVLEP NRYETFEPEYIVTGHYKRRDIEDLEVTERLINCKEDIHVPLDWTIPKAFGRF AKLAATFCGSA"

CDS

9391..12852
/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"
/codon_start=1

THIS PAGE BLANK (OPTIONAL)

THIS PAGE BLANK (OPTIONAL)

```

13245..14177
/gene="ipp2"
13245..14177
/gene="ipp2"
/note="YM6156_09, IPP2 gene (PPA2), len: 310, CAI: 0.14,
SM: PP2_YEAST, P28239, inorganic pyrophosphatase,
mitochondrial, contains P800387 Inorganic pyrophosphatase
signature"
/codon_start=1
/product="Ipp2p"
/protein_id="CA89230.1"
/db_xref="GI:809090"
/db_xref="SWISS-PROT:P28239"
/translation="MTLRNNATLSKARSRLEKQKTLNLSIRNRQSTIQGSKKTT
GKKKTLILNGEVSFFHVPDLNDEHETVMILVEVRMTTGKFEISKELRNPVIT

```

OY	1	tgaaactgtatgaagcaagaagatgtgcaagcatcaagcttcttaagatattcca	60
Db	27785	TGAAGTGTGATGACAAAGAGAGAGCTGCACACACATCAAAGCTTTTACGATGATTTCCA	27726
OY	61	ttcaaaaactctgtgaatcatgaaaccaagttcaacaagaattttcttgaattaaatgagt	120
Db	27725	TTCAAAAAATTTGTGATATGAAACCAAGTTCACCAAGAAATTTTCTTGATTTAAATAGATT	27666
OY	121	atataatagaataggagagacccgtatggccaagaagacgaattacaagattttcttc	180
Db	27665	ATATATATGCAATAGGGGAGACCCGTATGGCCAAAGACAGAAATTTACAGATTTTATAC	27606
OY	181	ttcaacagctgtcttaacgatatcgaaagcaattctgcaacttgaagttagttgcacaact	240
Db	27605	TTCAACAGCTGTTTAAGATATCGAAAGCATTTGCACTTGAAAGTTAGTGTACAAACTT	27548
OY	241	atcccaacatttttaaaaaacaagtcagttccctccaaacaactcaagaagcattatc	300
Db	27545	ATCCAAACATTTTAAAAAACAAGCTCAGTACCCCTTAAAAACAACATCAAACGCCATTATC	27486
OY	301	tgaagattcaatatctgcacacaggtaaagttatctatcgtgcgccagttccagtcatt	360
Db	27485	TGAGGATTCAATATGCCACACAGGTAAAGGTATCTATGTCGCCCAAGTTCAAGCTCATT	27426
OY	361	aacgcacagtaactcttcatccaagaagtcagttatttttaactagcgtatgcgtctac	420
Db	27425	AACGCCAGTAACCTTCTTCATCTCCAGAGGTATGTTATTTTAACTACGAGTACGCTAC	27366
OY	421	ttccctcgaaatttaccagaccagattttcaataaigtgttagattgctccctttaagta	480
Db	27365	TTCCCTCGAAATTTACGACACAGATTGTTCATTAATAGGATTTGGATTTGCTCCCTTAAAGTA	27306

QY	481	caaaagaaacctaataatgaaacttggaaatttaaaagaacctaagaacttaattcc	540
Db	27305	CAAAAGAAACCTTAATATGGAACCTTGGAAATTTAAAGACACTTAAGAACTTTAAATTC	27246
QY	541	aagtttgaagccgacctatgttgaagttatcgttcgttcgaagtaatgaatttcga	600
Db	27245	AAGTTTGAAGCCGTACGTATAGGTTTAAATGGGTTGCACTAATGATTAATTTGA	27186
QY	601	aaacacattcgggttagcgaagaattgataccactattctgttagcgaatgacaata	660
Db	27185	AAACACATCTGGCGTACGGAAGATGTGATCCCAATTTCTGTGGCAAGTACAAATA	27126
QY	661	aaaaaaccttgaaaaatttcgttcaattcttcttatagatatagatatagttgatt	720
Db	27125	AAAAACCTTTGAAAAAATTTCTGTTACTTTTCTTAATAGATATAGATATATGATG	27066
QY	721	gcttataatgaagttattatcgcgcctcttgtattccacttatataaattcttt	780
Db	27065	GCTTATAGATGAAGTAAATTAATTCGCGTCTTGATATCCCATTAATTAATAATTC	27006
QY	781	taaaatgatttctcgtgtgccttttgttgccttcgtatttttttttttggaccactg	840
Db	27005	TAAATGCAATTTCTCGTGCCTTTGTTGCTGATTTTATTTTATTTTGGACCACTG	26946
QY	841	gatggaaaaccttgatgatttttctactatttctaagttaactaataatcgagatt	900
Db	26945	GATGGAAAACCTTGATGATTTTATACCTTATTTTAAGTAAATTAATATCGAGATT	26886
QY	901	cagagacaacaatagaattcttcttcgtcagaataataacgcaataatctatgctt	960
Db	26885	CAGGACAAACATATGAATTTCTTTTGCAAGAAAAATTAACGAATTAATTTGATGCTT	26826
QY	961	tgaactactgactgtctgtcatagagagaacacagacaatgctacaaattcacaact	1020
Db	26825	TGACTACTGACTGCTGTGCATAGAGAGAACCAAGCAATGCTACATCTCAACTT	26766
QY	1021	tctagtctcctatacatctccatactctatgtctacgcgtacatacgaacttcaacc	1080
Db	26765	TCTAGTCTCCTATACATCTCCATCTCATATGCTACCGCTAATGCACTTACCAAC	26706
QY	1081	agaagggctaaacagagataactccctcgaagataaactctgcagtcgaataactacat	1140
Db	26705	AGAAAGGCTTAAGACAGATTAATCTCTGACAGATTAATCTGCAATGCAAAACCTTACAT	26646
QY	1141	tatgctaaattcccaacccttttttggcgcccttttcaacatccaatatcttaagcac	1200
Db	26645	TATGCTAATTTCCAAACCTTTTGGTGCCCTTTTACAATCTCAATATCTAGTACAC	26586
QY	1201	ttcacataaagttccaagacgaacttttagccttaagttatcccaagttatacgc	1260
Db	26585	TTCAATATTAAGTTTCAAGGACGCAATTTTGGCTTAGGTATATCCACGCTTATTAACG	26526
QY	1261	tgcacttgcacaaccttggacattcagccagtctgtgaagaacttaacgaaatgattgc	1320
Db	26525	TGCATTGCCAAACCTTGGCAATTCAGCCACTTGTGTAAACCTTAACGAATATGTGTC	26466
QY	1321	gatgttatcgaacctatattgtggaccggttttagattttgatatatcatatttaaa	1380
Db	26465	GATGTTATTGACCTATATATTTGGACCGGTTTGAATTTGTATATATCAATTAATAA	26406
QY	1381	tccaaagagcctatactatgaagatttttccatgatattccctgaatatattvgagttcag	1440
Db	26405	TCCAAAGAGCCTATATCTTGAAGATTTTACACGAATTTCTGTAATTTTGGAGCTTCAG	26346
QY	1441	gaatttatatttcacacaataaactagagagaatatatttaacgcgtcaatgccttgcac	1500
Db	26345	GAAATTTTATTTGACCAAAATTACTGAAGAAATTTTACACGTAAATGCTTTGACTAC	26286
QY	1501	gtactttaaacttaacccgattcgcacaactaagatcatcaacagttatttggcaaccatc	1560
Db	26285	GTACTTAACCTAATACCGCAATTCGCAACTAAGCATACAAAGTATTTTGGCAACATC	26226
QY	1561	gcttttttttggacttgcgcagcaacacatgcttataagcaattaaagaagatcccat	1620

THIS PAGE BLANK (U.SPTO)

THIS PAGE BLANK (U.SPTO)

Db 26225 GCTTTTGGACTTGGCAGCAGACACCATGCTTATGACCAATTAAGAAAGGCTCCAT 26166
 QY 1621 gacaactgttccatctcgtcgcacaatgtcccaattttacacaacacttttgg 1680
 Db 26165 GACAACGTGTTCCATCTGCTGACACACATGCTTCAAAATTTTATACCAACATTTTGG 26106
 QY 1681 agggtaacaaattgtatctgtaagaacagcgcggaacattgtgtcgcataaccc 1740
 Db 26105 AGGGTTAAACCAAGTTGTATTCGTAACACAGCGGCAACCTAAGGCGCTGCAATACCT 26046
 QY 1741 gactgccccttgcaatacctatggttccctgctccttcaagattgaattacattcac 1800
 Db 26045 GCATGCCCTTTCGAATATCAAGGGGCTTCCGGCTCTTCAAGATGGAATTTACATTTTAC 25986
 QY 1801 agtagtagacaagaagaacgtgagcattcccaattgtctccaatcgggaatagtcta 1860
 Db 25985 AGTAGTAGACAGAAAGAGTGGACCATTTCCAAATTTGCTCAATCTGGAATTAAGTGTCTA 25926
 QY 1861 ctctgcactgctgtccttgatataatccctgaagaatccctacaaactctgtatag 1920
 Db 25925 CTTCGCACTGCTGTGCTTGGATTATATCCCTGAAGATACCTTACAACTCTGTAGG 25866
 QY 1921 aaccctgttataagaataaacccttagcctttttacgtactgtataccgtttaaat 1980
 Db 25865 AACTCCTGTATAGAAATACCCCTTTAGCCTTTTACGTAAGTATACGTTAAAT 25806
 QY 1981 ttccatgtactataacattttttcactactataatgaattcttcgagcagccggtct 2040
 Db 25805 TTCCATGTATACATTAACCTTTTTCACATTTATGAAATCTATCGACGACCGGCT 25746
 QY 2041 ttgttgcagaagatgaaataacgtgttctgtgttggttgtaagaatttggagac 2100
 Db 25745 TTGTGTACCGAAGAGTGAATAAATCAAGTTTGTGTTGTAAGAAATTTGGAGAC 25666
 QY 2101 tctaaagttccatctactgttatacgaactacatacaagctcgtctgtatgtat 2160
 Db 25685 TATAAAGTACCTTACTTTGTTATACGACTCAATACAGTGGTCTGAGTGTAT 25626
 QY 2161 tgaagttgcagatcctaaagtagtagagaaggtgtgcataatagtttcgacgttttc 2220
 Db 25625 TGAAGTTGCAGATCTAAGTAGTAGAAGAGGCGCATTAAGTTTCGAGGTTTTC 25566
 QY 2221 ttttttcaaggtttcattgtgtctcctgaatttaaggtccttagttgtgtgt 2280
 Db 25565 TTTTAAAGGTTTATTTGTTGCTCTCAAAATTTAAGGCTTATGTTGTTGTTG 25506
 QY 2281 ttgtgtgttactatattttcaattcaagaaggaatttagtctgtcttataatgtccat 2340
 Db 25505 TTGTGGGTTACATATTTTCAATTCAGAGGAATTTAGCGTCTTTAATAAGTCCAT 25446
 QY 2341 agagataacgagacatgtctgtactacatacgaagcagataagcgatcgctagtcaag 2400
 Db 25445 AGAGATAACGAGACATGCTGCGTACTACATCAAGCATTAAGCGATCGGTAGCAAGG 25386
 QY 2401 gataaacggaagctgaagttttatgtctgtcagacagtccttgcaatgtcaatccgagc 2460
 Db 25385 GATTAACGGAAGTCTGAAGTTTGTATGCTGCAAGTCCCTTGAATGAATCCGACG 25326
 QY 2461 gtaaaaaacctaanaaatgtcgattgtgtcgaatgtgattactattgtatccagaatta 2520
 Db 25325 GTAAAAAACCTTAAGAATTTGCTGATTGGTCAATGATTACTTATGATCCGAATTA 25266
 QY 2521 gataaaatcgttggggaatctagtgtggaagacatcatgtgtcgcagacatccagt 2580
 Db 25265 GATTAATAATCGGTGGGGAATCTAAGTAGGAGACATCATGTGCGACACATCCAGT 25206
 QY 2581 tctgcgtcaatgcgaatgtacacacacccgtttaatacacagataagatccact 2640
 Db 25205 TCTGCGCTCAATGCGAAGTACACACACCGCTTATTAATACACGATATAGCGATCAACT 25146
 QY 2641 ccgctagaagacttgcagggggggttaactcaggagataagactcctcaataagactaa 2700

Db 25145 CCGTAGAGAACTTGCAATGGGAGGGGTAACCTAGGATGAATCTCCATTAAGCTAAA 25086
 QY 2701 caagtaactactatggtataaaaaagtttcaactcctccacgaagaatatagtct 2760
 Db 25085 CAAGTAACCTACTATGATATAAAAAAGGTGTTCTCTCTCCATCCAGGAATTAATAGT 25026
 QY 2761 aacgtattaaagaactattatcgtgttcccgcaatcaaacacttaagct 2820
 Db 25025 AACGTTTAAAGAAAACCTTATATAGGTTCCCGCAATCAACCCCTAAGTTAGCCT 24966
 QY 2821 gataattcctagagctgttacaagaatacttatacaaatatacaactaagcgaatagt 2880
 Db 24965 GATAATTTCTAGAGCTTGTACAAAGTACTTAAACAAATATACAACTAAGCGCAATGCT 24906
 QY 2881 gaagataatgatgggaatcagcaatgaataacgaatattgaggaataagggaataa 2940
 Db 24905 GAAGATAAATGANGGGAATAGCATGAATAATGAATATGAGATATGAGGAGGATAAA 24846
 QY 2941 gaatcaca 2948
 Db 24845 GAATCACA 24838

RESULT 2

LOCUS D50276 4156 bp DNA PLN 10-FEB-1999
 DEFINITION Saccharomyces cerevisiae Hst1 gene, complete cds.
 ACCESSION D50276
 VERSION D50276.1 GI:1777313
 KEYWORDS HST1, high-copy suppressor of TPP sensitivity.
 SOURCE Saccharomyces cerevisiae (strain: RAY-3a) DNA.
 ORGANISM Saccharomyces cerevisiae
 Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE

1 (bases 1 to 4156)
 Tsuchiya, E.
 Direct Submission
 Submitted (13-APR-1995) to the DDBJ/EMBL/GenBank databases. Eiko
 Tsuchiya, Hiroshima University Faculty of Engineering, Department
 of Fermentation Technology, Kagami Yama, Higashi-Hiroshima,
 Hiroshima 739, Japan (E-mail: etsuchi@ipc.hiroshima-u.ac.jp,
 Tel: 0824-24-7868, Fax: 0824-24-7868).
 2 (bases 1 to 4156)
 Tsuchiya, E., Matsuzaki, G., Tsukao, A. and Miyakawa, T.
 Molecular cloning and characterization of S. cerevisiae HST1 gene
 unpublished (1995)

REFERENCE

Tsuchiya, E., Matsuzaki, G., Kurano, K., Fukuchi, T., Tsukao, A. and
 Miyakawa, T.
 The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance
 to high concentration of Ca²⁺ with the participation of
 HST1/NRC1/BFR1
 Gene 176 (1-2), 35-38 (1996)

TITLE

JOURNAL
 MEDLINE
 FEATURES
 source

gene

CDS
 standard name="high-copy suppressor of TPP sensitivity"
 /codon_start=1
 /protein_id="BAA08819.1"
 /db_xref="GI:1616450"
 /translation="MSNRDNESMLRTSSDKAIAISQRDKREKSEVILIAASLDNEIRSY
 KIKRLSIGSDLLIDPELDIFEGESGSRSGTSSASAPSDTTVNNTRYSDP
 PLENLHGRNGSIESNKTGQNGYIGIKGVSPERKLANAVLAKNLMLVPAHQHPN
 VAPDNLTEVQPTIONIOISDNGEDNDGNSNENNDLEDNGESQSYENKKNNTNL
 NGLSRHNASLIRRSRTLRSTYEDPDNDKDSASEYVNRKYEERISIKKEEPV
 SIRDITTEELTKISNAGLTNDNDITLARIILSNAGSISDKKDPQPPGHGDEGIGST

location

Location/Qualifiers
 1..4156
 /organism="Saccharomyces cerevisiae"
 /strain="RAY-3a"
 /db_xref="taxon:4932"
 946..3693
 /gene="HST1"
 946..3693
 /gene="HST1"
 /standard name="high-copy suppressor of TPP sensitivity"
 /codon_start=1
 /protein_id="BAA08819.1"
 /db_xref="GI:1616450"
 /translation="MSNRDNESMLRTSSDKAIAISQRDKREKSEVILIAASLDNEIRSY
 KIKRLSIGSDLLIDPELDIFEGESGSRSGTSSASAPSDTTVNNTRYSDP
 PLENLHGRNGSIESNKTGQNGYIGIKGVSPERKLANAVLAKNLMLVPAHQHPN
 VAPDNLTEVQPTIONIOISDNGEDNDGNSNENNDLEDNGESQSYENKKNNTNL
 NGLSRHNASLIRRSRTLRSTYEDPDNDKDSASEYVNRKYEERISIKKEEPV
 SIRDITTEELTKISNAGLTNDNDITLARIILSNAGSISDKKDPQPPGHGDEGIGST

THIS PAGE BLANK (1.3PT6)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: us-09-165-460a-2

Perfect score: 2342

Sequence: 1 MEDLKITLIDHPNIPKLIIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BL0SUM62
Gapop 10.0, Capext 0.5

Searched: 100059 segs, 36664827 residues

100059

otal number of hits satisfying chosen parameters:

100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2336	99.7	453	ST24_YEAST	P47154 saccharomyc
2	903	38.6	474	ST24_SCHPO	010071 schizosacch
3	698.5	29.8	475	ENCL_HUMAN	075844 homo sapien
4	328	14.0	426	YHEN_BACSU	P40769 bacillus su
5	161	6.9	286	HTPX_PASWJ	P57846 pasteurella
6	150.5	6.4	310	HTPX_HELPJ	Q9ZK84 helicobacte
7	150	6.3	335	HTPX_ARCFU	030004 archaeoglob
8	147.5	6.4	283	HTPX_HAEIN	P44840 haemophilus
9	142.5	6.1	310	HTPX_HELPY	025882 helicobacte
10	139	5.9	366	HTPX_AERPE	Q9Y667 aeropyrum p
11	138.5	5.9	258	HTPX_METTH	026669 methanobact
12	130.5	5.6	292	HTPX_PYRHO	058997 pyrococcus
13	130.5	5.6	317	HTPX_THEAC	Q9H1V2 thermoplas
14	127.5	5.4	2136	YCF2_MARPO	P09875 marchantia
15	120	5.1	292	HTPX_BUCAL	P57406 buchnera ap
16	119.5	5.1	2294	YCF2_ARATH	P56786 arabidopsis
17	119	5.1	297	HTPX_STRGC	Q30795 streptococc
18	116	5.0	293	HTPX_ECOLI	P23894 escherichia
19	113	4.8	284	HTPX_METSU	Q59076 methanococc
20	113	4.8	298	HTPX_BACST	Q31657 bacillus su
21	109	4.7	506	YCX2_EUGGR	P31821 euglena gra
22	109	4.7	593	YCB6_METUA	Q58688 methanococc
23	107.5	4.6	1032	MT18_YEAST	P40469 saccharomyc
24	106.5	4.5	646	NTPL_HAEPU	Q37319 heliothis a
25	106.5	4.5	1029	YFC5_YEAST	P43571 saccharomyc
26	105.5	4.5	873	COX1_ACACA	Q37370 acanthamoeb
27	105	4.5	1427	SRB8_YEAST	P25648 saccharomyc
28	104.5	4.5	585	YHD9_YEAST	P38732 saccharomyc
29	104.5	4.5	1100	LOSI_YEAST	P33418 saccharomyc
30	104	4.4	302	HTPX_AOUAE	Q67798 aquifex aeo
31	104	4.4	530	CP14_CHICK	P79690 gallus gall
32	103.5	4.4	185	GU01_RAT	P35894 rattus norv
33	103.5	4.4	450	VATH_SCHPO	014265 schizosacch

ALIGNMENTS

RESULT	1	ST24_YEAST	STANDARD:	PRT:	453 AA.
AC	P47154	ST24_YEAST	STANDARD:	PRT:	453 AA.
DT	01-FEB-1996	(Rel. 33, Created)			
DT	01-FEB-1996	(Rel. 33, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC ENDOPEPTIDASE 1) (PPEP 1) (A-FACTOR CONVERTING ENZYME).				
GN	STE24 OR AEC1 OR YJR117H OR J2032.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
NCBI_TaxID=4932;					
SEQUENCE FROM N.A. AND FUNCTION.					
Rose M., Koetter P., Entian K.D.;					
submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.					
SEQUENCE FROM N.A. AND FUNCTION.					
MEDLINE=97167681; PubMed=9015299;					
Fujimura-Kamada K., Nouvet F.J., Michaelis S.;					
"A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor."					
J. Cell Biol. 136:271-285(1997).					
FUNCTION.					
MEDLINE=97218305; PubMed=9065405;					
Boyarctchuk V.L., Ashby M.N., Rine J.;					
"Modulation of Ras and a-factor function by carboxyl-terminal proteolysis."					
Science 275:1796-1800(1997).					
FUNCTION.					
MEDLINE=98365461; PubMed=9700155;					
Tam A., Nouvet F.J., Fujimura-Kamada K., Slunt H., Sisodia S.S.,					
Michaelis S.;					
"Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAAX processing."					
J. Cell Biol. 142:635-649(1998).					
FUNCTION.					
MEDLINE=98393572; PubMed=9725832;					
Boyarctchuk V.L., Rine J.;					
"Roles of prenyl protein proteases in maturation of Saccharomyces cerevisiae a-factor."					
Genetics 150:95-101(1998).					
FUNCTION.					
MEDLINE=20158928; PubMed=10692417;					
Schmidt W.K., Tam A., Michaelis S.;					
"Reconstitution of the Ste24p-dependent N-terminal proteolytic step in yeast a-factor biogenesis."					
J. Biol. Chem. 275:6227-6233(2000).					
[7]					

34	103	4.4	645	1	VP74_NPVAC	P15963 autographa
35	103	4.4	1101	1	KOES_YEAST	Q08217 saccharomyc
36	577	4.4	577	1	THH1_SCHPO	Q09684 schizosacch
37	102	4.4	2493	1	YBA4_YEAST	P3194 saccharomyc
38	100.5	4.3	260	1	YMI6_RECAM	Q21266 recitmonia
39	100.5	4.3	1040	1	RIK1_SCHPO	Q10426 schizosacch
40	100	4.3	1325	1	MRP4_HUMAN	Q15439 homo sapien
41	99.5	4.2	309	1	Q7AA_HUMAN	Q76100 homo sapien
42	99	4.2	485	1	YC11_KLEBN	Q48457 klebsiella
43	99	4.2	1024	1	RPOB_PLAFA	P21421 plasmidium
44	98.5	4.2	313	1	NO2M_RHISA	Q99817 rhiplephal
45	98.5	4.2	417	1	O65A_DROME	P82982 drosophila

RP CHARACTERIZATION. PubMed-10825201;
 RX MEDLINE-20285442; PubMed-10825201;
 RA Trueblood C.E., Boyarchuk V.L., Picooglou E.A., Rozema D.,
 RA Poulter C.D., Rine J.,
 RT "The Caax proteases, Aicp and Rcep, have overlapping but distinct
 RT substrate specificities."
 RL Mol. Cell. Biol. 20:4381-4392(2000).
 RN [8]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE-98409630; PubMed-9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.,
 RT "Endoplasmic reticulum membrane localization of Rcep and Ste24p,
 RT yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal a-factor cleavage."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1175-1180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR MATING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249617; CAAB9647.1; -
 DR EMBL: U77137; AAB38271.1; -
 DR MEROPS: M48.001; -
 DR SCD: S0003878; STE24.
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase: Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum; Pheromone response.
 FT TRANSMEM 13
 FT TRANSMEM 33
 FT TRANSMEM 90 110
 FT TRANSMEM 122 142
 FT TRANSMEM 168 188
 FT TRANSMEM 198 218
 FT TRANSMEM 307 327
 FT TRANSMEM 358 378
 FT METAL 297 297
 FT ACT_SITE 298 298
 FT METAL 301 301
 FT METAL 390 390
 FT ACT_SITE 394 394
 FT SEQUENCE 453 AA; 52324 MW; 331CC9AE2D7C99DA CRC64;

Query Match 99.7%; Score 2336; DB 1; Length 453;
 Best Local Similarity 99.8%; Pred. No. 3.7e-160;
 Matches 453; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFDTLLDHPNIPWKLIIISGFSIAQSFESYLYROYOKLSEKLPVLEDEIDDETFH 60
 Db 1 MFDTLLDHPNIPWKLIIISGFSIAQSFESYLYROYOKLSEKLPVLEDEIDDETFH 60
 QY 61 KSRNYSAPAKKFSIFGDVYMLAOKLVFIKYDLPKIMHMAVSLNAVLPAVFHNVSTVAQ 120
 Db 61 KSRNYSAPAKKFSIFGDVYMLAOKLVFIKYDLPKIMHMAVSLNAVLPAVFHNVSTVAQ 120
 QY 121 SLCEFLGLSLSTLVDPPLSYSHFVLEKRGFKNLVQMLITDMISLTLAVAGPIL 180
 Db 121 SLCEFLGLSLSTLVDPPLSYSHFVLEKRGFKNLVQMLITDMISLTLAVAGPIL 180

QY 181 YLFKIDKPPPTDPLKIMFVFLVQILLAMTITVFIPMPKNTPTLEDGELKSIESTA 240
 Db 181 YLFKIDKPPPTDPLKIMFVFLVQILLAMTITVFIPMPKNTPTLEDGELKSIESTA 240
 QY 241 DRVGFPPDKIFIDGSRSSSHSNAYFGLPPTSRIYFDVLVNSNDELTAVLAHIG 300
 Db 241 DRVGFPPDKIFIDGSRSSSHSNAYFGLPPTSRIYFDVLVNSNDELTAVLAHIG 300
 QY 301 HMOKNHIVNVYFISQHTFLFSLFTSIYRNTSFYNTFGFLEKSTGSPVDPVITKEPPI 360
 Db 301 HMOKNHIVNVYFISQHTFLFSLFTSIYRNTSFYNTFGFLEKSTGSPVDPVITKEPPI 360
 QY 361 IIGFMFLNDLLTPLECAMQVMSLISRTHEYQADAVAKKGYKNLCRALIDQIKNST 420
 Db 361 IIGFMFLNDLLTPLECAMQVMSLISRTHEYQADAVAKKGYKNLCRALIDQIKNST 420
 QY 421 MNVDPLSSYHSHPTLAERSTADYSEKKKN 453
 Db 421 MNVDPLSSYHSHPTLAERSTADYSEKKKN 453

RESULT 2
 ST24_SCHPO STANDARD; PRT; 474 AA.
 ID ST24_SCHPO
 AC Q10071;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-
 DE SPECIFIC ENDOPROTEASE 1) (PPSEP 1).
 GN SPAC3H1.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN-972;
 RA Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED PROTEINS (BY SIMILARITY).
 CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z68144; CAA92258.1; -
 DR MEROPS: M48.001; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum.
 FT TRANSMEM 103 123
 FT TRANSMEM 196 216
 FT TRANSMEM 230 250
 FT TRANSMEM 344 364
 FT TRANSMEM 381 401
 FT METAL 332 332
 FT ACT_SITE 333 333
 FT METAL 336 336
 FT SEQUENCE 336 AA; 336 MW; 336 MW; 336 MW;

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

6512.739 Million cell updates/sec

Title:	US-09-165-460A-1
Perfect score:	1825

Sequence: 1 acctacctttttctatct.....aaataaacgattaacatt 1825

Scoring table: . IDENTITY_NUC

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Maximum DB seq length: 200000000000

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_un:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_ov:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htgo_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
C	1	1823.4	99.9	1825	8	SCJCR117W	Z49617 S. cerevisiae	Z49617 S. cerevisiae
C	2	1553.4	85.1	1706	8	SC077137	U77137 Saccharomyces	U77137 Saccharomyces
C	3	342	18.7	1555	8	SCJCR116W	Z49616 S. cerevisiae	Z49616 S. cerevisiae
C	4	195	10.7	99360	8	NCEB11N2	AF515344 Neurospora	AF515344 Neurospora
C	5	187	10.2	36920	8	SPAC33H1	Z68144 S. pombe	Z68144 S. pombe
C	6	137.2	7.5	1275	8	AF35372	AF35372 Arabidopsis	AF35372 Arabidopsis
C	7	136	7.5	686	6	AB62270	AB6270 Sequence 92	AB6270 Sequence 92
C	8	136	7.5	686	6	ARI55763	ARI55763 Sequence	ARI55763 Sequence
C	9	136	7.5	686	6	E66288	E66288 Genome DNA	E66288 Genome DNA
C	10	121	6.6	935	8	SCJCR118C	Z49618 S. cerevisiae	Z49618 S. cerevisiae
C	11	117.4	6.4	1859	9	AF064867	AF064867 Homo sapiens	AF064867 Homo sapiens
C	12	117.4	6.4	2963	9	AB016068	AB016068 Homo sapiens	AB016068 Homo sapiens
C	13	117.4	6.4	2966	9	HS113834	E13834 Homo sapiens	E13834 Homo sapiens
C	14	117.4	6.4	2968	6	E32056	E32056 Human AFCL	E32056 Human AFCL
C	15	114.2	6.3	1340	3	AF358443	AF358443 Physarum	AF358443 Physarum
C	16	98.4	5.4	78592	3	AC004335	AC004335 Drosophila	AC004335 Drosophila
C	17	98.4	5.4	102413	2	ACG20024	ACG20024 Drosophila	ACG20024 Drosophila
C	18	92.8	5.1	41327	3	CBRG47719	CBR4676 Caenorhabditis	CBR4676 Caenorhabditis
C	19	91.6	5.0	40767	3	CCE04412	Z81461 Caenorhabditis	Z81461 Caenorhabditis
C	20	88	4.8	317511	1	CJ1168X3	AL113906 Campylobacter	AL113906 Campylobacter
C	21	86.6	4.7	582	8	CNS018P3	AL110644 Botrytis	AL110644 Botrytis
C	22	81.6	4.5	660	8	CNS018H3	AL112937 Botrytis	AL112937 Botrytis
C	23	72.2	4.0	261	9	AK027884	AK027874 Homo sapiens	AK027874 Homo sapiens
C	24	66.8	3.7	1072	6	AX055334	AX055322 Sequence	AX055322 Sequence
C	25	66.6	3.6	124808	2	AP000652	AP000652 Homo sapiens	AP000652 Homo sapiens
C	26	65.6	3.6	16488	1	AEO04992	AEO04992 Halobacter	AEO04992 Halobacter
C	27	65	3.6	256172	2	AC005139	AC005139 Plasmodium	AC005139 Plasmodium
C	28	64.4	3.5	169433	2	AC069309	AC069309 Mus musculus	AC069309 Mus musculus
C	29	63.4	3.5	186306	2	AC083946	AC083946 Mus musculus	AC083946 Mus musculus
C	30	63.6	3.5	1245	6	AX055202	AX055202 Sequence	AX055202 Sequence
C	31	63.6	3.5	19715	1	AEO00555	AEO00555 Helicobacter	AEO00555 Helicobacter
C	32	63.6	3.5	234112	3	PFMAL374	AE001347 Plasmodium	AE001347 Plasmodium
C	33	62.4	3.4	13383	3	AE001374	AE001374 Plasmodium	AE001374 Plasmodium
C	34	61.8	3.4	98734	4	PFMAL1P2	AE011745 Plasmodium	AE011745 Plasmodium
C	35	61.6	3.4	14943	1	AEO01582	AE001582 Helicobacter	AE001582 Helicobacter
C	36	59.8	3.3	12029	3	AEO01392	AEO01392 Plasmodium	AEO01392 Plasmodium
C	37	59.8	3.3	110000	2	AC087895_0	AC087895 Mus musculus	AC087895 Mus musculus
C	38	59.2	3.2	114736	3	PFMAL333	Z298547 Plasmodium	Z298547 Plasmodium
C	39	59.2	3.2	253305	3	PFMAL37	AL034559 Plasmodium	AL034559 Plasmodium
C	40	58.8	3.2	98899	2	AL603743	AL603743 Danto rer	AL603743 Danto rer
C	41	58.8	3.2	318221	1	PFMAL13P3	AL049184 Plasmodium	AL049184 Plasmodium
C	42	58.8	3.2	318503	2	AC074166	AC074166 Mus musculus	AC074166 Mus musculus
C	43	58.4	3.2	172239	3	AC069587	AC069587 Homo sapiens	AC069587 Homo sapiens
C	44	58.2	3.2	12029	3	AEO01412	AEO01412 Plasmodium	AEO01412 Plasmodium
C	45	58.2	3.2	187469	9	AL357312	AL357312 Human DNA	AL357312 Human DNA

ALIGNMENTS

RESULT 1					
SCYRI17W					
LOCUS	SCYRI17W	1825 bp	DNA	PLN	11-AUG-1997
DEFINITION	S.cerevisiae chromosome X reading frame ORF YCR117W.				

VERSION: 249617.1 GI:1015836

ORGANISM

REFERENCE 1 (bases 1 to 1825)

REFERENCE 2 (bases 1 to 1825)

JOURNAL

Submitted (25-SIP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

FEATURES
source
Martinsried, FRG; E-mail: Neues@mlps.emblnet.org

Location/Qualifiers
1. 1825

/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="X"

gene

CDS

/gene="STRE24"
343..1704
/gene="STRE24"
/note="ORF YJRI17w"

/codon_start=1
/protein_id="CA89647.1"

/db_xref="GI:1015837"

/translation="MDKTIIDHPNIPKRIISGSIASFESLYRQXKSET
KIPVLEIDIDETFRKSNYSKAKRKEIFGDYVLAQKVFIRKDEPKIRHAYS
LINAIVPVRHVSVAQSLCEGLSSSTLDVLSYSHVELEKGFNKLIVOL
WITDKISLTAVIAGPILYFLKIFDEPDELMYINVEFVVOILMTIIPVIM
PNEKFTPLEDELEKRSISLADRVGFPDKIFVIDGSRSSNSNYPGLPFSKRI
VLEQTVNSNDIEPAVLAHEIGHOKNHYVMVIFESOLHFLISLFTSIKNSF
VTFGPELEKSGSFVDVITKEPPIIIGPMLENDLPLECOMOVMSLISRTHYO
ADAYAKKLCYKONLCNALIDLQIKNLSTNANVPLYSYHSPTLAEKLTADYSEK
KKN"

BASE COUNT 570 a 397 c 280 g 578 t
ORIGIN

Query Match 99.9%; Score 1823.4; DB 8; Length 1825;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1824; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 acctactttttctatcttcaacagaaagccttaacacacacacataactc 60
DB 1 acctactttttctatcttcaacagaaagccttaacacacacacataactc 60
OY 61 acatacatacaacatacatatataatgaactgtatattcctataacaaa 120
DB 61 acatacatacaacatacatatataatgaactgtatattcctataacaaa 120
OY 121 aagaagcaatlaaactttccctctttctcagtcattactcaaaaacttaattc 180
DB 121 aagaagcaatlaaactttccctctttctcagtcattactcaaaaacttaattc 180
OY 121 aagagcaatttaactttccctctttctcagtcattactcaaaaacttaattc 180
DB 121 aagagcaatttaactttccctctttctcagtcattactcaaaaacttaattc 180
OY 181 tctctctgttctgcctttttccgaaaaaaatcgacgggaataaaaaaaagac 240
DB 181 tctctctgttctgcctttttccgaaaaaaatcgacgggaataaaaaaaagac 240
OY 241 aacgaacaagaagaagctcgcaactataaacacttcataatcaagaagaaga 300
DB 241 aacgaacaagaagaagctcgcaactataaacacttcataatcaagaagaaga 300
OY 301 aggaaaaaaagaagaagaataagaactgcagccttatcatgtttgaatcaagc 360
DB 301 aggaaaaaaagaagaagaataagaactgcagccttatcatgtttgaatcaagc 360
OY 361 attctgacacatacttaataccggtggaataatcaattcttggttctcgattgccaa 420
DB 361 attctgacacatacttaataccggtggaataatcaattcttggttctcgattgccaa 420
OY 421 tttctctgaaacttaactgaagctacagagatcaagaagcattctgaaaaaagt 480
DB 421 tttctctgaaacttaactgaagctacagagatcaagaagcattctgaaaaaagt 480
OY 481 ccacctgtctgaagaagaagaatgcatgaataactttcataatcaagaagaactacc 540
DB 481 ccacctgtctgaagaagaagaatgcatgaataactttcataatcaagaagaactacc 540
OY 541 cgggccaagccaagcttcacattctcgtagctctataacctaagccaagaagctgt 600
DB 541 cgggccaagccaagcttcacattctcgtagctctataacctaagccaagaagctgt 600
OY 661 ttcatacaatacagacactcttcctcaaaatctgcacatgycgcgtttcttattgaatgca 660

DB 601 ttcatacaatacagacactcttcctcaaaatctgcacatgycgcgtttcttattgaatgca 660
OY 661 gtcctgcagctgaagatttataatgctcaccatgtctgcacagattatgcttcgggt 720
DB 661 gtcctgcagctgaagatttataatgctcaccatgtctgcacagattatgcttcgggt 720
OY 721 ccttaacagctgtctcaccatgtctgttgaattgccaactctctacataagcattgtc 780
DB 721 ccttaacagctgtctcaccatgtctgttgaattgccaactctctacataagcattgtc 780
OY 781 ctggaagaagaatcttggtttcaataaattgacccgtccacatgatacagatatgac 840
DB 781 ctggaagaagaatcttggtttcaataaattgacccgtccacatgatacagatatgac 840
OY 841 aagagctgacttttggtgcatctatggtgcccacatccttaactgtctcctaagac 900
DB 841 aagagctgacttttggtgcatctatggtgcccacatccttaactgtctcctaagac 900
OY 901 ttgataaaatcccccctcctcattcctttgtacatattgtctcttctgtctgtccaa 960
DB 901 ttgataaaatcccccctcctcattcctttgtacatattgtctcttctgtctgtccaa 960
OY 961 attctagccaatgaacatcctcagctctcaatgacatgcccagtgcttaataagtcactca 1020
DB 961 attctagccaatgaacatcctcagctctcaatgacatgcccagtgcttaataagtcactca 1020
OY 1021 ttggagagcgtggaactgaaaaaatctatgaaatttggtgcgatagtgtgttcct 1080
DB 1021 ttggagagcgtggaactgaaaaaatctatgaaatttggtgcgatagtgtgttcct 1080
OY 1081 ctgataagaattttgtcaattgacggctcaaaaagctcttcattcaaaagcatatttc 1140
DB 1081 ctgataagaattttgtcaattgacggctcaaaaagctcttcattcaaaagcatatttc 1140
OY 1141 acaggtttgcattccaccccaagaagaattgtttctgcacacttaagtaagaagtaat 1200
DB 1141 acaggtttgcattccaccccaagaagaattgtttctgcacacttaagtaagaagtaat 1200
OY 1201 tctactgataaattgaagcgttttgccacatgaatccgttcacatgycgaagaagaacac 1260
DB 1201 tctactgataaattgaagcgttttgccacatgaatccgttcacatgycgaagaagaacac 1260
OY 1261 atcgttaataatgctatcttgatgaatgaagaacttcctacttttccctttacc 1320
DB 1261 atcgttaataatgctatcttgatgaatgaagaacttcctacttttccctttacc 1320
OY 1321 agcatctacagaatacatcatatttcaaacacacttcgcttttcttagaagaagtcact 1380
DB 1321 agcatctacagaatacatcatatttcaaacacacttcgcttttcttagaagaagtcact 1380
OY 1381 ggcagttctgtgaatccggttaccactaagaatccccaatcatgtgaattatgta 1440
DB 1381 ggcagttctgtgaatccggttaccactaagaatccccaatcatgtgaattatgta 1440
OY 1441 ttttaagacttataactccacacagatgagcatgcaattctgtgaattatattcc 1500
DB 1441 ttttaagacttataactccacacagatgagcatgcaattctgtgaattatattcc 1500
OY 1501 agaacctatgaataatcaagctgactgaatgcaaaaatctgggctcaagaagaactca 1560
DB 1501 agaacctatgaataatcaagctgactgaatgcaaaaatctgggctcaagaagaactca 1560
OY 1561 tgaaggctcctaattgatctacaatacaaaactttccacatgaattgataatcctctg 1620
DB 1561 tgaaggctcctaattgatctacaatacaaaactttccacatgaattgataatcctctg 1620
OY 1621 tattctagctatattatcccatccaactctagctgaagaatcgacgcttagactat 1680
DB 1621 tattctagctatattatcccatccaactctagctgaagaatcgacgcttagactat 1680
OY 1681 gttagtgaagaagaagaactaaatctatagtagtacatattagcatgtacggttaact 1740

QY	331	aggccttatctatcgttgaatctttaagagcattctccgacacccctaatatcccgctggaaa	390
QY	331	aggccttatctatcgttgaatctttaagagcattctccgacacccctaatatcccgctggaaa	390
Db	61	AGGCGTTTATCTACGTTGGATCTTAAGAGGATCTCGACCAATCTCAATATCCCTGGAAA	120
QY	391	ttaatcatcttcgggtgtcttcgcatctggcccaatttcttcgcgaatcttaactgaatagaa	450
Db	121	TTAATCATTTCTGGGTCTTCGATTCGCCCAATTTCTTTCGAATCTTACTTACGTACAGA	180
QY	451	cagttaccagaagcgtatctgaaacaaagtctgcaccctgtgcctgaaagcgaatgtgat	510
Db	181	CAGTACACAGAGCTATCTGAACAAAGTGGCCACTGCTGGGAAGAGCAATGTGATGAT	240
QY	511	gaacattctcataaatccaaggaacctaactccggcgccaaaggccaagttctcatttggct	570
Db	241	GAACCTTTTCATTAACAAGGAACACTCTCCGGGCAAGGGCCAAATTCCTCATTTTGGCT	300
QY	571	gaagctcataacctaagcccaaaagcagttttctacccaataagacaccttccccaatc	630
Db	301	GAGGTCTATACCTTACCCCAAAAGCTAGTTTTCATCAATAGACCTTCTCCCTTAATTC	360
QY	631	tggcacatggcggcttctcttaatgtaatgcagtcgcctgcagtcagaattcataatgtctcc	690
Db	361	TGGGACATGGCGCTTCTTATTTGAATGCAAGTCAGTCGACAGTTCATATGCTGCC	420
QY	691	actgtgcacagaagttatgtctcttggtctctctataccagttgtcttaactgtgtgat	750
Db	421	ACTGTGCACAGAGTTTAAAGCTCTTGGGGTCTCTTATCCAGTTGTCTACCTTGGTGTAT	480
QY	751	ttggcaactctctacatatagccattttgtccctggaaagaaaaattggtttcataaatgt	810
Db	481	TTGGCACTCTCTTACATATAGCCATTTTGTGCTCGGAAGAAAATTTGGTTCAATTAATGTG	540
QY	811	accgtccaataatgataccacgaatgatcatcaagagctcgaacttggctatgacatgtgt	870
Db	541	ACCGTCCAATATAGATATCAACCATATAGATCAAGAGTCTGACTTGGCTATGCTATGTGT	600
QY	871	ggcccaactccttaactcgttcccttaagatccttgaataatccctactgaattcccttgg	930
Db	601	GGCCCAATCTTACCTGTTCTCTTAAGATCTTGAATTCCTTAAGATTCCTTAAGATTCCTTGG	660
QY	931	taactatgtctctcttgttctgtgtgcacaaacttaagccatgaacataatccacgtctc	990
Db	661	TACATTAATGATCTTCTTGTGCTGTGTCACAAATCTTAAGCATATGATTCACATCTTC	720
QY	991	atcatgcacatgtttaaataagattcactcccatcgtgaagacggtgaaactgaaaaactct	1050
Db	721	ATCATGCCCATGTTTAATTAAGTCACTCCATTGGAGAGACGGTGAACTGAAAAAATCTATT	780
QY	1051	gaaaagtgtgcgatagaagttggtgtccctctagataaagatttgttaatgtagcgctca	1110
Db	781	GAAAGTTTGGCCGATATAGATTTGGGCTTCCCTTAAGATAAGATTTTGTCAATGACGGCTCA	840
QY	1111	aaaaagattcttcataccaacgcgatatattccaaggtttgcattcaactccaagaagaatt	1170
Db	841	AAAAGATCTTCTCATTTCAACCGCATATTTTCAACAGGTTTCCATTCCTCCCAAGAGATT	900
QY	1171	gtttgtgtcgaactttagtgaacgagtaattctacgatgaattaaagcgtgtttgtgcc	1230
Db	901	GTTTGTGTGCACACTTTAAGTGAACATTAATTCTACTGATGAATAATACGGCTGTTTGGCC	960
QY	1231	catgaatcgtgtcactgtgcaaaaaaacacacatcgttaataatgtaatcatttagtcaatg	1290
Db	961	CATGAATATCGGTACTACGGCAAAAAAACCAACATCGTAAATATGTCATCTTGTGCAATTG	1020
QY	1291	caaccccttctaatttctccctttccaccagatcttaagaataacatcatatttacaac	1350
Db	1021	CACACCTTCTCCATTTTCTCCCTTTTCCACACAGCATCTACAGAAATACATCAATTTTACAC	1080
QY	1351	acctgtgcgttttctctagaagaagtcacactgtgcgagtttggtagcccgattacaaga	1410
Db	1081	ACCTTGGCGCTTTTCTTAAGAAAGTCCACTGCGAGTTTGTGTAATCCCGTTATCTACTAAG	1140
QY	1411	gaattcccatcatcatgtgattatgtaatttaagacttaattaaactcaactcgaagat	1470

THIS PAGE BLANK (1.3PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds
(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1
Perfect score: 1825
Sequence: 1 acctacattttttctatct.....aaataacgattaacatt 1825

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_cm:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hcgo_hum:*
31: em_hcgo_inv:*
32: em_hcgo_rnd:*
33: em_hcg_hum:*
34: em_hcg_inv:*
35: em_hcg_rnd:*
36: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1823.4	99.9	1825	8	SCYR117W
2	1553.4	85.1	1706	8	SCYR7137
3	342	18.7	1555	8	SCYR116W
4	195	10.7	99360	8	NCB11N2
5	187	10.2	36920	8	SPAC3H1
6	137.2	7.5	1275	8	AF353722
7	136	7.5	686	6	AB6270
8	136	7.5	686	6	AR155763
9	136	7.5	686	6	B66288
10	121	6.6	935	8	SCYR118C
11	117.4	6.4	1859	9	AF064867
12	117.4	6.4	2963	9	AB016068
13	117.4	6.4	2966	6	HSY13834
14	117.4	6.4	2966	6	B32056
15	114.2	6.3	1340	3	AF358443
16	98.4	5.4	78592	3	AC004335
17	98.4	5.4	102413	2	AC020024
18	92.8	5.1	41327	3	CBRG47319
19	91.6	5.0	40767	3	CEC04F12
20	88	4.8	317511	1	CU11168X3
21	86.6	4.7	582	8	CNS018P
22	81.6	4.5	660	9	AK027874
23	72.2	4.0	2161	6	AX055334
24	66.8	3.7	1072	6	AX055334
25	66.6	3.6	124808	2	AP000652
26	65.6	3.6	16488	1	AE004992
27	65	3.6	256172	2	AC005139
28	64.4	3.5	169423	2	AC069309
29	64.4	3.5	186306	2	AC083946
30	63.6	3.5	1245	6	AX055202
31	63.6	3.5	19715	1	AE000555
32	63.6	3.5	234112	3	PFMAL4P2
33	62.4	3.4	13383	3	AE001374
34	61.8	3.4	98734	2	PFMAL1P2
35	61.6	3.4	14943	1	AE001528
36	59.8	3.3	12029	3	AE001392
37	59.8	3.3	110000	2	AC087899.0
38	59.2	3.2	11736	3	PFMAL3P3
39	59.2	3.2	253305	3	PFMAL3P7
40	58.8	3.2	98899	2	AL603743
41	58.8	3.2	318221	2	PFMAL1P3
42	58.8	3.2	318503	2	AC074166
43	58.4	3.2	172239	2	AC069587
44	58.2	3.2	12029	3	AE001412
45	58.2	3.2	187469	9	AL357312

ALIGNMENTS

RESULT 1
SCYR117W 1825 bp DNA 11-AUG-1997
LOCUS SCYR117W
DEFINITION S.cerevisiae chromosome X reading frame ORF YR117W.
ACCESSION Z49617.1
VERSION Z49617.1 GI:1015836
KEYWORDS
SOURCE
ORGANISM
baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 1825)
AUTHORS Rose, M., Koelter, P. and Entlian, K.D.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1825)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopfersplatz 18a D-82152

QY 271 aaaccactctataattacaggaagaagaaaaaaagyggaatagaaaactgc 330
|||||
|||
Db 1 AAACCACTTCATTAATTACAGGAAAAGGAAGAAAAAAGGAGAATATGCAAAACTGC 60

QY	331	agcccttatcaatglttgatctctaaagactctccgaacatcccaatcccaatccgltgaaa	390
Db	61	AGGCCCTTAAATCAATGTTTGATCTTAAAGACGATTCTCGACACATCTTAATATCCCTGTGAAA	120
QY	391	ttaactaattcttggtcttcgatattgcccgaattctcttccgaactctaaactgaagtacaga	450
Db	121	TTAATCAATTTCTGGGTTCTCGATTGCCCCAATTTTCTTTCGAAATCTTACTTACGTACAGA	180
QY	451	cagtaacagaagactatctgtaaacaaaagtgtgccaactgtgcttgyaagacgaaatltgatat	510
Db	181	CAGTACACGAAGCATATCTGAAACAAAGTTGCCACCTGTGCTGGAAGACGAATATGATGAT	240
QY	511	gaaactttcataataatcaagaagaactactccggcgccaaggtcttccatttgggt	570
Db	241	GAACCTTTTATTAATATCAATCAAGAACTACTCCGGGGCCAAAGCCAAATTTCTCATTTTGGGT	300
QY	571	gaactctataactgaagcccaaaagactagttttcataaataagaaacttcccctaaatc	630
Db	301	GACGCTTAATACTTAGCCCCAAAAGCTAGTTTTCATCAATAGACACTTTCCTCTAANAATC	360
QY	631	tggcacatggccggttcttcattatgaaatgacgtccgcgcagtaagaattcataatgctcc	690
Db	361	TGGCACAATGGCCGTTTCTTATTATGAATGCAAGTCGCGCCAGTCAGATTTCAATATGCTGCC	420
QY	691	acgtctgcacagagtttaagctctctggtctcttaaccagttgtctaaccttggtgat	750
Db	421	ACTGTCCGACAGAGATTTATATCTTCTTGGGCTCTTATCCAGATTTCTACCTTGATGAT	480
QY	751	ttggcaactctcttaactatgaacatttgcctcggaagaaaatttggtttaataaattg	810
Db	481	TTGGCACTCTTCTAATAATACCAATTTTGCTCGGAAGAAAATTTGGTTTCAATAAATG	540
QY	811	accgtccaactatgataccagatatacatcaagaagctcacttggcgtatgcatttgt	870
Db	541	ACCGTCCAAGTATGAGATCCACCGATATGATCAAGAGCTCACTTGGCGATATGATTTGCT	600
QY	871	ggcccaatcccttaaccctgtctcctaagaatcttgataaatccctactgattcccttgg	930
Db	601	GGCCCAATCCCTTACCTGTTCCTTAAAGATCTTGAATAATTCCTACATGATTTCCCTTGG	660
QY	931	taactatgactctctgtctcgttgcacaaactctgaagatgaagaataatccagcttc	990
Db	661	TACATTAAGGCTCTCTGTGTCGTTGTCACAAATTTAGCCATATCAATATTTCCAGCTTC	720
QY	991	atcatgccaactgtttaaataagttcacctccaatgaaagacggtgaaactgaaaatctcat	1050
Db	721	ATCATGCCCCATGTTAATTAATGTTCACTCCATTGAGAGACGGTGAACATAAANAATCTATT	780
QY	1051	gaaagtttggccgagatagatttgggtctccctctagataaagaatttgcatttgaagcgtca	1110
Db	781	GAAAGTTTGGCCGATAGAGATGTTGGGTTCCCTCTGTGATTAATATTTTGTCAATGACGGCTCA	840
QY	1111	aaaagactcttcatatcaaacgcatcttccaaggtttgcacattcaactccaagagatt	1170
Db	841	AAAAGACTTCTCTCAATTAACAGCATTTTTCACAGGTTGCCATTTACCTCCAAAGAAATTT	900
QY	1171	gttttgttgcagaaactttagtgaacagtaaatctctacgtatgaaatctagcgtcttggcc	1230
Db	901	GTTTGTGTCGACACTTTAGTGAACAGTAATTTACTAGATGAATTAACGGCTGTTTGGCC	960
QY	1231	catgaaactcgttcatcttggcaaaaaaacacacatcgtataatagtatcactttagtcaattg	1290
Db	961	CATGAATTCGGTCACTGTGCAAAAAAACACATCTGTTAAATATGTCATCTTTAGTCAATTG	1020
QY	1291	caacacttccacatcttctcccttccacagagatctaaagaataatacatctttaaacac	1350
Db	1021	CACACCTTCTCTCATTTTCTCTCTTTCCACACAGCATTTACAGAAATACATCTTTTACAC	1080
QY	1351	aaccttgcgcttcttcttagaagaagctcacgtgcagtttctgtatcccglttataactaag	1410
Db	1081	ACCTTCGGCTTTTCTTAAAGAAAGTCACAGGAGTTTGTGATCCGGTATATCACTAATAG	1140
QY	1411	gaattcccacatcatatgattatgtaatttaacgactatctaactcccaactcgaatggt	1470

QY	331	aggcccttatcatgtttgatctccttaagagatcttcagcaatcccaatcccaatccgctggaaa	390
Db	61	aggcctttatcactagtttgatctttaaagacgatttcgcacatcccaatattccctggaaa	120
QY	391	ttaaatcattctctggcttcgcagcttgcgaattctcttcgaatcttaactgaagtacga	450
Db	121	tttatcatcttctgggtcttcgcantggcccaattttcttgcgaattcttactgactacaga	180
QY	451	cagttacagaagctatctgaaacaaagtgtgcacctgtgtcgtgaagacgaatatgatat	510
Db	181	cagttacagaagctatctgaaacaaagtgtgcacctgtgtcgtgaagacgaatatgatat	240
QY	511	gaaactttcatataccaaggaaactctccggcgccaaggccaagtttccatttctgt	570
Db	241	gaaacttttatataccaaggaaactctccggcgccaaggccaagtttccatttctgt	300
QY	571	gagctcataactagcccaaaagctagtttccatcaaatagacacttccctcaaatc	630
Db	301	gagctctataactagcccaaaagctagtttccatcaaatagacacttccctcaaatc	360
QY	631	tggcacatggcgtttctctatgtaaatgcagtcctgtcgaatcagaattcatatgctctcc	690
Db	361	tggcacatggcgtttctctatgtaaatgcagtcctgtcgaatcagaattcatatgctctcc	420
QY	691	actctgcacaaggatttatgctctctgggtctctatccagttgtctcaacttggttat	750
Db	421	actctgcacaaggatttatgctctctgggtctctatccagttgtctcaacttggttat	480
QY	751	ttgcacactccttaactaaaccaattctgtcccggaagaaaattcgtgttcaataaatc	810
Db	481	ttgcacactccttaactaaaccaattctgtcccggaagaaaattcgtgttcaataaatg	540
QY	811	accgtccaaactatgatcacacgatatgatcaaaagtctgaatttggcgtatgtatgt	870
Db	541	accgtccaaactatgatcacacgatatgatcaaaagtctgaatttggcgtatgtatgt	600
QY	871	ggcccaatccttactactcgtttcccttaagactcttgaataaaattccctacagatttcccttgg	930
Db	601	ggcccaatccttactactcgtttcccttaagactcttgaataaaattccctacagatttcccttgg	660
QY	931	tacattatgtctctctgttctgtctgtgcacaaactcttagacatgaacatccacgtcttc	990
Db	661	tacattatgtctctctgttctgtctgtgcacaaactcttagacatgaacatccacgtcttc	720
QY	991	atcatgcccatlgtttaataaagtctacatccattgtagagacggtgaaactgaaaaaatctat	1050
Db	721	atcatgcccatlgtttaataaagtctacatccattgtagagacggtgaaactgaaaaaatctat	780
QY	1051	gaaagtttggcgcgatagagtgtgtgtccctctagataaagatttctgtcatatgaagctca	1110
Db	781	gaaagtttggcgcgatagagtgtgtgtccctctagataaagatttctgtcatatgaagctca	840
QY	1111	aaaagatcttctcatcaaaaacgcatatttcaagaaggttgcacttaactccaagaagaatt	1170
Db	841	aaaagatcttctcatcaaaaacgcatatttcaagaaggttgcacttaactccaagaagaatt	900
QY	1171	gttttgttcgacactttagtgaacaglaattctactagatgaataatcagtcgttttggcc	1230
Db	901	gttttgttcgacactttagtgaacaglaattctactagatgaataatcagtcgttttggcc	960
QY	1231	catgaataatcgttcaactgtggcaaaaaaacacatcgtataatagtcatctttagtcaattg	1290
Db	961	catgaataatcgttcaactgtggcaaaaaaacacatcgtataatagtcatctttagtcaattg	1020
QY	1291	cagaccttccatatttctcccttccaccagacatcaagaataatcaatatttaaac	1350
Db	1021	cagaccttccatatttctcccttccaccagacatcaagaataatcaatatttaaac	1080
QY	1351	accttgcgcttttcttaagaagaatccacatcgtgagtttgtgtacccgttatcactaag	1410
Db	1081	accttgcgcttttcttaagaagaatccacatcgtgagtttgtgtacccgttatcactaag	1140
QY	1411	gaattcccatlcatcaltgattatgtatttaaacagatataactccactcgaatgt	1470

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460A-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaag.....ggggaggaataaagaatcaca 2948

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched:

1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters:

2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rtd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rtd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	2948	100.0	29987	8 SC8156	249260 S.cerevisia
2	1560.4	52.9	4156	8 D50276	50276 Saccharomyc
3	1361	46.2	4130	8 YSCNCR1G	142821 Saccharomyc
4	1163	39.5	4225	8 SCU65682	165682 Saccharomyc
5	1069.4	36.3	3250	8 SCU22580	032580 Saccharomyc
6	931	31.6	3525	8 SCUD3849	053849 Saccharomyc
7	786.8	26.7	3240	8 YSCBU1	050083 Yeast BUL1
8	703.8	23.9	3020	8 YSCGNC58X	X8901 S.cerevisia
9	725.2	7.6	270	8 YSCGNC58X	X87379 Yeast Eco R
10	143.0	4.9	27559	8 SC8339	249210 S.cerevisia
C 11	75	2.5	2033	8 SCYNR068C	271683 S.cerevisia
12	71.8	2.4	4775	8 SCU32938	032938 Saccharomyc
C 13	68	2.3	780	11 CNS06K8B	AL402561 T7 end of
14	65.6	2.2	256172	2 AC005139	AC005139 Plasmodiu
15	65.6	2.2	310779	2 AC005140	AC005140 Plasmodiu
16	64.4	2.2	904	11 CNS06189	AL399967 T3 end of
17	64.4	2.2	86827	3 PFMA13P5	AL034556 Plasmodiu
18	63.4	2.2	1920	3 PFGRUB	X62393 P.falciparu
19	63.4	2.2	110000	2 PFMA13P2_0	AL049185 Plasmodiu
20	62	2.1	253305	3 PFMA13P7	AL034559 Plasmodiu
C 21	61.6	2.1	1141	6 AX083744	AX083744 Sequence
22	60.8	2.1	224112	3 PFMA14P2	AL035475 Plasmodiu
23	60.6	2.1	202872	2 AC016160	AC016160 Homo sapi
C 24	60	2.0	5632	3 AF010892	AF010892 Dictyoste
C 25	59.4	2.0	3214	3 AF019980	AF019980 Dictyoste
C 26	59.4	2.0	318221	2 PFMA13P3	AL049184 Plasmodiu
27	58.8	2.0	321003	2 PFMA14P3	AL035476 Plasmodiu
28	58.4	2.0	78439	17 HSU85197	U85197 Homo sapien
C 29	58.4	2.0	169546	2 AC004157	AC004157 Plasmodiu
30	58.4	2.0	245802	2 AC006279	AC006279 Plasmodiu
31	58.4	2.0	251124	9 HUAE000660	AE000660 Homo sapi
C 32	58.2	2.0	153098	3 PFMA13P2	AL034558 Plasmodiu
C 33	58	2.0	169794	2 AC004688	AC004688 Plasmodiu
34	58	2.0	178137	9 AC011302	AC011302 Homo sapi
C 35	57.8	2.0	13684	3 AE001403	AE001403 Plasmodiu
36	57.8	2.0	153477	2 AC006278	AC006278 Plasmodiu
37	57.6	2.0	53932	2 AC023371	AE0023371 Homo sapi
C 38	57	1.9	12029	3 AE001431	AE001431 Plasmodiu
C 39	57	1.9	163443	2 AC006280	AC006280 Plasmodiu
C 40	57	1.9	178273	2 AC005308	AC005308 Plasmodiu
C 41	57	1.9	196149	2 AC004709	AC004709 Plasmodiu
C 42	56.8	1.9	67970	3 PFMA11P3	AL031746 Plasmodiu
C 43	56.8	1.9	149627	9 AC087428	AC087428 Homo sapi
C 44	56.6	1.9	92633	2 PFMA14P1_3	Continuation (4 of
C 45	56.6	1.9	153477	2 AC006278	AC006278 Plasmodiu

ALIGNMENTS

RESULT	1
SC8156/c	29987 bp DNA
LOCUS	SC8156 29987 bp DNA
DEFINITION	S.cerevisiae chromosome XIII cosmid 8156.
ACCESSION	249260 271257
VERSION	249260.1 GI:809081
KEYWORDS	BULL1; cytochrome b5; DNGL; initiation factor 1A; inorganic pyrophosphatase; IIP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF1; TPS3; transfer RNA-Gln; trehalase-phosphate synthase; U6 snRNP; URA10; baker's yeast.
SOURCE	Saccharomyces cerevisiae
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE	1 (bases 1 to 29987)
AUTHORS	Iye,G., and Churcher,C.M.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 29987)
AUTHORS	Barrell,B. and Rajandream,M.A.
TITLE	Direct Submission

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1HQ E-mail: barrell@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, embl entry SC9920, accession no. Z48639 and at the end by cosmid 8021.

FEATURES

1. 29987
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8156"
/map="13R"
CDS
complement(<1..356)
/note="Ym8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends Ym9920.13c"
/codon_start=1
/product="unknown"
/protein_id="CA89242.1"
/db_xref="GI:809082"
/db_xref="SWISS-PROT:003496"
/translation="MTDVOFSSDEIEFRVKEFLIANNPANNENDAVLQIEHD
FRYLQYKGLPLNNESTRILPDTFSICLRSHQIIASKRIDSQEFLSAVKEQLT
EANNIIEYVLDEFWA"
misc-feature
1. 140
/note="overlap with SC9920 Z48639 S.cerevisiae chromosome XIII cosmid 9920"
gene
complement(851..1312)
/gene="TIF11"
CDS
complement(851..1312)
/gene="TIF11"
/note="Ym8156.02c TIF11 gene, IFLA_YEAST, P3891, Initiation factor 1a"
/codon_start=1
/product="Tif1p"
/protein_id="CA89243.1"
/db_xref="GI:809083"
/db_xref="SWISS-PROT:P38912"
/translation="WGKNTGKGRGKRGKNDSPKRELIYKEGQEVATITKMLGN
GRVESCDFGKRNHRIKRLKRYMGGODIILVSLDFDDQCDVYHKYMLDEART
LKNQELPENAKINETDNEGFESDEDVNFEGNADEDEGEDELIDDI"
gene
complement(1678..4842)
/gene="TPS3"
CDS
complement(1678..4842)
/gene="TPS3"
/note="Ym8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-trehalose-phosphate synthase and PIR:S27471, PS00213 Lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="WTIIVASILFYPTQFQFADVTNSTAKLVESMAIKVDONNOELS
NNKQSSSTVSASGHYIGLQEQAINCEQIRANVGPATGVNHNHNMESLQFL
EELANATHANAGCIPANNVPVSGSTAPORVVEFFSAPARVCSPOEASASISL
SRSSAHNDLSSSLKKNPNLSPDSHPPRVRSKSAVITPVSKSPVDPAVDVAKY
REFEQOQASLPKRVSGSTAGDSSIASSSSLRLRSQFQDFIDTDSSEDDIDSLDLE
TDATKYNVPFGGYSNNAKRLASLMSRYELFKLPTIIVSDKNGSLNNAVNA
AEKTVKEPVSVMGTGIPITDELPHVECHKISIKLEQDFSSFPVVTDDITFGAYKVA

misc-feature

CDS

CDS

CDS

CDS

CDS

KOLLPTLHYIIPNPNKAEDEHMDYQYVQKQFSDRIYSVYRPGDIWTHYHLM
LVPQVREKILPRKAGFLVHVSFSSVEFRLANERILEGILGNFVGQFKEKRR
FLQCNRLADVSDNDEVKYCNIVSWAPIGIDYHILTSQLRNGSYLSEPOLKER
WRNKLILYCRDQPDVIRBGLQCKMLAYEFLIENPYLEKVVYIOICIGSSPEYER
IMVVDRLNSNSISISOPVPLHODIDFPOYALINEADVELDALREGNNLKEH
FIVSFEENAPLLISEPSSSYLKEGAILPNPHINNYAOSIKRSLDESPERKRRH
KKLRKSVIEHSDWMTIKCFEYINMANSNOETVFNWADPEKFCADKSKHLEIF
KISEPFSRLMLSLSESSNNIVYLSSEFTNTEFSLNGVNLGILANGVYVRNG
SWYIVIEEDDMKKEVAKIFDEKVERLPGSYKRIADSMTRFENSDDDOVRYPYGEA
ITHINTLEDDDHIAVYKRDIVFYQDYLAAAEYIKRFLKFGVSGVSPDMSRSLSR
SSSVGKNNKHFONOVDFVCVOSTSPILPEFLVQVEKNNLKKGYTILYSSSR
STVAKEHNGVNEFTLHDTAA"
complement(2260..2301)
/gene="TPS3"
/note="PS00213 Lipocalin signature"
5199..6140
/note="Ym8156.04, unknown, len: 313, CAI: 0.11, previously
noted ORF, SW:YTP3_YEAST P38430"
/codon_start=1
/product="unknown"
/protein_id="CA89245.1"
/db_xref="GI:809085"
/db_xref="SWISS-PROT:P38430"
/translation="MKLIYDAHCYVITDPDPTFCGDCGSGGTLRCVASSNPDYNNL
KILAGRSTSKNDICVGFVHPVSHLEFVGSRRKRVSHYQVLYEKNQEDSLVOVL
PEPDLLEYIKRENDLVSYIGIDGRLPLANGVYMNENKRLITVYKLSHOE
TVFRFRWLARHTSKPISIHVRCCHKDILYVNGVAKILCHSYTGSKETLLG
OVLKKPPDRIFVSLSKINKDEDEGALVRSIPSTILNETVDPIPNPDGSYKAL
TEQOLYNIAARAMDETLDSQAALRYENFQFIK"
6392..6997
/note="Ym8156.05, unknown, len: 301, CAI: 0.12,
SW:YTP2_YEAST P38429 previously noted ORF"
/codon_start=1
/product="unknown"
/protein_id="CA89246.1"
/db_xref="GI:809086"
/db_xref="SWISS-PROT:P38429"
/translation="MARPVNNAETESGRPTGGGVASNNNGSCNNNGSNNNNNN
NNNNNSNNNSNNNGPTSSGRTNGKORITAAQOQYIKLITHTIDNHPDLRPSHPM
DFEETVAFLEIRYKDHFDVLDVNLDTLOGYLGSLGSLAKTYSYKRNNGOGDHRKHKR
DLANVVRHHPDEHSIKETDCIPRTIYKKNKKKFKMEFRG"
7278..7889
/note="Ym8156.06, unknown, len: 301, CAI: 0.13,
SW:YTP1_YEAST P38428, extends previously noted ORF"
/codon_start=1
/product="unknown"
/protein_id="CA89247.1"
/db_xref="GI:809087"
/db_xref="SWISS-PROT:P38428"
/translation="MEDSRLLITLILYGVYFLKFFQSNQHPNSAQRISANGVNAHR
PGSTONALRRGTGVNGHPVTTQMVETVOVALNPLHBOIRYSLENTGSVEYERY
LRGDFSPPEPPEPSRAPMGNAALVDNNAAGGEPDRKRMICAEMLDKFHDYLN
EDMSNLSFKDIDIERKRLIYQARKNLERLQSDKQSLIT"
complement(8013..9398)
/note="Ym8156.07c, unknown, len: 461, CAI: 0.11"
/codon_start=1
/product="unknown"
/protein_id="CA89248.1"
/db_xref="GI:809088"
/db_xref="SWISS-PROT:Q03508"
/translation="MEEFEEFRKRGKSGNHRVLRKNSACCELAVPEVEPHAI
TKLIYDENDTLFATPGTBOLEYRELLNLTSLTPNGMNEGFLQAAIEISIKR
PRRSNNAIDIVKLAESYSADDTISYLTGSESKFHLIHAQOTASNNKCSNER
RPNVYLCARRAISERTSKRELKADPLEYIPSLSELSITDRKHQIDAFSFFSISL
LPRKWSAIVPPTKSLPREQLEWMDVARKNNSQALSTQFDLRRTPROQIYIL
CTASRLHLSLEVILYKLRKRGRTFRRLYHPYIPCAEFGKDIIPALEIKAVSND
TRPDSKKTVOHISOIYLLBOPGACIYHFOVTLAVIASHHNRKRPLEYEFKAPEP
NRYEITPEPEYIYGHYKKNRIDLEVEYERLINCKEDIMWPLDNTIPIKAFGRF
AKLAATFCNSN"
9991..12852
/note="Ym8156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"
/codon_start=1

```

/note="YH8156.09, IPP2 gene (PPA2), len: 310, CAT: 0.14,
SW:IP22_YEAST, P28239, Inorganic pyrophosphatase,
mitochondrial; contains P500387 Inorganic pyrophosphatase
signature"
/codon_start=1
/product="IPP2p"
/protein_id="CA809250.1"
/db_xref="GI:809090"
/db_xref="SWISS-PROT:P28239"
/trslation="MULTIMULTICARS1EPLKQTLNLSIRHROESTIQGSKYTT
GKATKLLINGVSGSEFDVPELDLNEHRTVNMVEVPRMTGKFELESKELRNPVIV

```

OY	1	tcgaactgttgatgacaaagaagatctcaacagctcaagaagttgttcgatgatttcca	60
Db	27785	TGAAGTGTGATGAAACCAAGAGAACCTACACAGCATCAAAAGTTGTGTGAGATTTTCCA	27726
OY	61	ttcaaaaatctgtgtgata.tgaacccaagtctcacagaagatttctcttgata.ttaaatgagct	120
Db	27725	TTCAAAATTTGTGATATGAAACCAAGTTCAACAAGATTTCTTGATTTAAATGAGTT	27666
OY	121	ataataatgaaatgaggggagacgctgaagccaaaggaaactgaaatttaacagatttattac	180
Db	27665	ATATAATTTGATATAGGGAGACCGTGAAGCCAAAGGAAGTGAATTAAGATTTATTATAC	27606
OY	181	ttcaacagctgtgttaacgatatcgaagaagatttgcacacttgaagttagtgttcacaactt	240
Db	27605	TTCAACAGCTGTTTAACGATATGGAAGATTTGGCAACTTGTGAAGTTAGTGTTCACAAGTT	27546
OY	241	atccaaacattttaaaaaaacaggtcagtaacccctaacaacactcaaaagcacgcatctc	300
Db	27545	ATCCAAACATTTTAAAAAACAGGTGATGATACCTTAAAAACACATCCAAAGCACAGCATTTATAC	27486
OY	301	ctgaagattccaatctgcgcacacaggtgaacgtagttcatcgcgtgcgcacgttcaagctaat	360
Db	27485	TGAGGATTCATATTCGCACACAGGTAAGGTTAGTTCAATGTCGCGCCAGTTCCAGCCTCAATT	27426
OY	361	aacgcacgataactcttcataccaagaagtagttatttcttaacc.tgaacggttagtcgctctac	420
Db	27425	AACGCCATTAATCTTCTTCATCCAAAGAGTAGTTATTATTTTAACTACCGGAGAGTCTGTAC	27366
OY	421	tttccttgaattttaaagccacgattgttcataaaagggttggagttgctgctcttcacgta	480
Db	27365	TTTCCTGAAATTTAAGACCAAGATTGTTCATAAAGGGTTAGGATTGCTCCTTTACAGTA	27306

OY	481	caacgagacacataatgtgcaacttggaaatttaataagaacattlaaggaacatttaattcc	540
Db	27305	CAACGAGACACTTAATGTGAACCTGGAAATTTAATAGACATTAGAGAACTTAAATTC	27246
OY	541	aagcttgaagaagcgcgaatgtgtgaagttctatctgcgtcgcgaagaatgaatlaattga	600
Db	27245	AAGTTTGAAGAAGCTGCCATGTGTAGGTTTATTCGCTTGAGTAATGATTAAATTTGA	27166
OY	601	aaacacatctggcgttagcgaagaatgtatatacccatcttcgttlaagcaagtcacaata	660
Db	27185	AAACCATCTTGGCGTAGGGAAGATGTGAATGCCCTTTCTGTTAGCAAGTACAAAAATA	27126
OY	661	aaaaaaccttgaaaaaaatttcgttactttttcttatagatatagatatatgtatggtt	720
Db	27125	AAAAACCTTGTGAAAAAATTCGTTACTTTTCTTAAGATTAATTAATATAGTATGTTT	27066
OY	721	gcttaataatgaagtatattatcggtcttccttgtatctcccatlaataataatcttt	780
Db	27065	GCTTATAGATGAAGTATTATTCGGCTCTTGATTCGCATATTTATATAAATTCCTT	27006
OY	781	taaaatgcatlcttggtgtccctttgttcttcctgtatcttttttttttttggacactg	840
Db	27005	TAAATGCATTTTCGGTGCCTTTGTGCTCTGTAATTTTATTTTGGACACATG	26946
OY	841	gatggaaaccttggatatttataccattatttaagtattcaaaaatcsgaatt	900
Db	26945	GATGGAATAACCTTGATGATTTTATTAACCTTATTTAAGTACTTAATAATCGAGATT	26886
OY	901	cagagacaaaacataagaattctctctgcgaagaaaaaataaacgnaataatgtatgctt	960
Db	26885	CAGGACAAAACATGTGAATTTCTTTGTGCAGAAAAATTAACGAAATTAATGTATGCTT	26826
OY	961	tgaactactgtctgtcatagagagaaacgaagaatgtctacaattctcaacatt	1020
Db	26825	TGACTACTGACTGTCTGTCAATAGAGAGAACCAAGACCAATGTCAATCTCAACTT	26766
OY	1021	tctagtgtcctatacatctccatacctatgtgttaccgctataatgcacattccaaac	1080
Db	26765	TCTAGTGTCTCTATACACTCTCCATATCCTATGTGTACCGCTAATATGACACTTCAAC	26706
OY	1081	agaagggctaaacagagataactccctcgaaagataaactctgcgaatgcaaaaacttacaat	1140
Db	26705	AGAAGGCTTAACAGAGATATTCCTCGAAGCATTAATCTCGCATGCAAAAACTTACAAT	26646
OY	1141	taatgtaattccaaaccttttttgggtcccttttttacaattccaaattactagtaaac	1200
Db	26645	TATGCTAATTTTCCAACCTTTTGTGTCCTTTTAACTCAATATCTAGTAACAC	26586
OY	1201	ttcaacataaagtttcaagagcgaatttttagcgttaagtatttaccgaagtataatgc	1260
Db	26585	TTCACATATTAAGTTTCAAGAGCGCAATTTTATAGGCTTAGGTATATCCACAGGTTATTACG	26526
OY	1261	tgcattgcgaacaccttggcaattcaagcgaatctgtctgaagaagacttaacgaagtgttc	1320
Db	26525	TGCATTTGCCAATCCTTGCGCAATTCAGCGCAAGTCTGTGAAGACTTAACGAATGTGTTC	26466
OY	1321	gacgttattgacctataatgtgtgaaccggttttagattttgtatataatcattataa	1380
Db	26465	GATGTTATTGACCTTAATTTGTGACCGGTTTTAGATTTGTATTAATATCAATTTATTTAA	26406
OY	1381	tcocaaagcgctatacttgaagaatttttacaatgaattctctgnaatatttggagttcag	1440
Db	26405	TCCAAAGCGCTAATACTTGAAGATTTTACCATCAATTCCTGAATTTATTTGGAGTTTCAG	26346
OY	1441	gaatttatatttgcacaaataacatgaagaaatatttacaagtcacatgcttttgaatc	1500
Db	26345	GAAATTTATTTTGGACCAATATACGTGAGAAATATTTTACACGTCAATGCTTTGACATAC	26286
OY	1501	gtacttaaacataatccgcatctgcaactaagatatacaaatgtatatttggcaacatc	1560
Db	26285	GTACTTAACCTTAATACCGCAATTCGCACTTAAGCATATCAAGTATATTTTGGCAACCATC	26226
OY	1561	gcttttttttggacttgcacgcacacacatgcttatagtagaatataacggaagtcacat	1620

Db 26225 GCTTTTTTGGACTTGGCCAGCACACCATCTTATGACATTAACAGAAAGCCCTCAT 26166
 Qy 1621 gacacgcttccatctgctgacacacatgcttccaattatatacacacacatttgg 1680
 Db 26165 GACAACTGTTCCATTGCTGACAAACATGCTTCCAAATTTTATACAAACATTTTGG 26106
 Qy 1681 agggtaacccaagtttgatcttgtaagaacagcgaggaaacctatggtctgcataact 1740
 Db 26105 AGGGTTAACCAAGTTTGAATTTGTAGAAACAGCGGGAACCTATGTTGCTGATATACCT 26046
 Qy 1741 gcaatcccttgcaatacatcaatgagggttccctgctcttcaaatgattgaatttacc 1800
 Db 26045 GCATCCCTTTCATATATCATGAGGCTTCTGCTTCTTCAAGATTTGATTAATTCATTTAC 25986
 Qy 1801 agtagtagacaagaagctgagcagcatctccaattgctcacaactggaataagtgcta 1860
 Db 25985 AGTAGTAGACAGAAAGCTGAGCAGCATTTCCAAATTTGCTCAATCTGGAATTAAGTCGA 25926
 Qy 1881 ctgcgactgctggtcccttgatataatccctgaagagataccttaacaaactctgtagg 1920
 Db 25925 CTTCGCACTGCTGGTCTTGATTAATTCCTGAAAGATTAACCTTAACAACCTCTGGTAGG 25866
 Qy 1921 aacitcctgtaataaataaccccttagcctttttagcttatacctttaaata 1980
 Db 25865 AACTCTGCTTATAGAAATTAACCTTTAGCTTTTACCTGATTAACCTTTAAAT 25806
 Qy 1981 ttccatgtaacataacatttcttcaactactatgtaattctatcgaagcagcagcgtc 2040
 Db 25805 TTCTATGACTATTAACCTTTTTCACATCTATTAAGAAATCTATCGAGCAGCAGCGGCT 25746
 Qy 2041 ctgtgtaaggaagtggaaaaaaacctgagatttctggtgcttggtagaagaatttgagac 2100
 Db 25745 TTTGTAGCGAAGAGTGAATAAATGAGTTTGTGTTGTTGTTGAAGAAATTTGAGAGAC 25686
 Qy 2101 cataagtaacataactctgtaataaccccttagcctttttagcttatacctttaaata 2160
 Db 25685 TATAAGTACTATTAATCTTTGTTTACGGAATTAACAGTCGTTCCGTTAGTGAT 25626
 Qy 2161 tgaagctgcaagatcgaagtagaagaagtgagcatctaaatagtttcgaagcttctc 2220
 Db 25625 TGAAGTGTACATCTAAGTAG 25566
 Qy 2221 tttcttaaggttttatttggctccctcagaatttaaggtctttagtttggctgt 2280
 Db 25565 TTTTGTAGGTTTATTTATTTGCTCTCAGAAATTAAGGCTTAAATTTTGTGTTGT 25506
 Qy 2281 ctgtggttcaatacttcaattcaataaggaatttagctgtcttcttaattatgccaat 2340
 Db 25505 TTTGGGCTTACATTTTTCATTTCAATTCAAAGAGAAATTTAGCTCTTTTATTAATGTCANT 25446
 Qy 2341 agagaatacgaagagcagctgctgtaatacatcaataagcgaataagcgaatcgaataag 2400
 Db 25445 AGAGATACGAGAGAGATCTGCTGCTACTACATCAAGCAGATTAAGCAGTCTGTCAGAAAG 25386
 Qy 2401 gataaaggaagctggaagtttggctgctgcacagctcccttgacaatgaaatcccgagc 2460
 Db 25385 GATAAACGGAAGTCTGAAGTTTGTGATCTCTGACAGTCCCTTGACAAATGAATCCGAGC 25326
 Qy 2461 gtaaaaaaacctaaagaattgctgcatggtggaatggaatttctattgattcagaata 2520
 Db 25325 GTAAAAAACCCTAAAGATTTGTCGATTGGCTTAATGATTTACTTATGATCCAGAAATTA 25266
 Qy 2521 gataataaactcgtggtggaatctagtggaagacatcatggtctgcgacgacatccagt 2580
 Db 25265 GATATTAATTCGCTGGGGAATCTAGTGGAGAGACATCATGCTCGGCAAGCATCCAGT 25206
 Qy 2581 tctgctgtaacgcaagtgcacacacacgctgtaataacacagatataagcgatccaaact 2640
 Db 25205 TCTGCTGTAATGCGAAGTGACACACACCGCTTATTAACACGATATAGCATCCAACT 25146
 Qy 2641 ccgctagaagaactgcatgggaggggttaactcagagatagaatctctcaataagactaa 2700

Db 25145 CCGTAGAGAACTTCATGCGGAGGCTAAGTACAGGATTAAGATTCCTCCATTAAGACTA 25086
 Qy 2701 caagtaactacttagtataaaaaagggtgtcactctccatccagaataataatgct 2760
 Db 25085 CAAGTAAGTACTTAGTATATAAAAAAGGTGTCACTGCTCCATCCAGGAAATTAATAGT 25026
 Qy 2761 aacgtattaaagaataactattatggttcccgccaatcaacaccttaagcttaagct 2820
 Db 25025 AACGTATTAAGAAAAAATTTATTAATGAGTTCCCGCAATCAACACCTTAACGTTAAGCT 24966
 Qy 2821 gataattccagaagctgttcacagaatacttatacaaatatacaactaagcagaatgct 2880
 Db 24965 GATAATTTCCTAGAGCTTGTATAGATATCTTCAAAATTAATCAACTAAGCAATGCT 24906
 Qy 2881 gaagataatgattgggaataagcaatgaataaataacgaatattgaggaataatggaggaata 2940
 Db 24905 GAAGATATATGATGGGAATAGCAATGAATAATACGATATTGAGGATTAATGGAGAGATA 24846
 Qy 2941 gaattaca 2948
 Db 24845 GAATCACA 24838

RESULT 2
 D50276
 LOCUS D50276 4156 bp DNA PLN 10-FEB-1999
 DEFINITION Saccharomyces cerevisiae HST1 gene, complete cds.
 ACCESSION D50276
 VERSION D50276.1 GI:1777313
 KEYWORDS HST1; high-copy suppressor of TFP sensitivity.
 SOURCE Saccharomyces cerevisiae (strain:RAY-3A) DNA.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 4156)
 AUTHORS Tsuchiya, E.
 TITLE Direct Submission
 JOURNAL Submitted (13-APR-1995) to the DDBJ/EMBL/Genbank databases. Eiko
 Tsuchiya, Hiroshima University Faculty of Engineering, Department
 of Fermentation Technology, Kagami-Yama, Higashi-Hiroshima,
 Hiroshima 739, Japan (E-mail:etsuchi@ipc.hiroshima-u.ac.jp,
 Tel:0824-24-7868, Fax:0824-24-7868).
 REFERENCE 2 (bases 1 to 4156)
 AUTHORS Tsuchiya, E., Matsuzaki, G., Tsukao, A. and Miyakawa, T.
 TITLE Molecular cloning and characterization of S. cerevisiae HST1 gene
 JOURNAL Unpublished (1995)
 REFERENCE 3 (sites)
 AUTHORS Tsuchiya, E., Matsuzaki, G., Kurano, K., Fukuchi, T., Tsukao, A. and
 Miyakawa, T.
 TITLE The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance
 to high concentration of Ca²⁺ with the participation of
 HST1/NRC1/BFR1
 JOURNAL Gene 176 (1-2), 35-38 (1996)
 MEDLINE 97075905
 FEATURES
 source
 CDS
 gene
 CDS
 Location/Qualifiers
 1..4156
 /organism="Saccharomyces cerevisiae"
 /strain="RAY-3A"
 /db_xref="taxon:4932"
 946..3693
 /gene="HST1"
 946..3693
 /gene="HST1"
 /gene="HST1"
 /standard_name="high-copy suppressor of TFP sensitivity"
 /codon_start=1
 /protein_id="BA08819.1"
 /db_xref="GI:1816450"
 /translation="MSNRDNEMLRTTSQKALASQDRKRSEVLLAQSLENEIRSY
 KNLKRLSIGMDLIDPELDIKFGESGRRSMGTSASSAMPSPDTTVNTRSPD
 TPLENHLGRNGIESNKGKGYLGIKRGVSPKRLNANVYKRLILWVPAQHPN
 VKPDNLEIWDPLNTOLDSNGEDNGNENNDNGEDKESQSYENKENTINL
 NRGISLRGNASLIRPSTLRSTTEPFDNEDDDKCGSASETVKVERISKIERPV
 SLRDIETELRKISNAGLTNDATLARTLSMASYSYDKRDQPPQEGHYDEGIGFST

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342
Sequence: 1 MEDLKTILDHPNIPWKLITS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2336	99.7	453	ST24_YEAST	P47154 saccharomyc
2	903	38.6	474	ST24_SCHPO	Q10071 schizosacch
3	698.5	29.8	475	FAC1_HUMAN	O75844 homo sapien
4	328	14.0	426	YHFN_BACSU	P40769 bacillus su
5	161	6.9	310	HTPX_PASMU	P57846 pasteurella
6	150.5	6.4	310	HTPX_HELPJ	O92K64 heliocobact
7	150	6.4	335	HTPX_ARCFU	O30004 archaeoglob
8	147.5	6.3	283	HTPX_HAEIN	P44840 haemophilus
9	142.5	6.1	310	HTPX_HELPY	O25582 helicobacte
10	139	5.9	366	HTPX_AERPE	O25582 aeropyrum p
11	138.5	5.9	258	HTPX_MERTH	O58669 methanobact
12	130.5	5.6	292	HTPX_PYRHO	O58997 pyrococcus
13	130.5	5.6	317	HTPX_THERAC	O2HJ22 thermoplasm
14	127.5	5.4	2136	YCF2_MARPO	P09975 marichantia
15	120	5.1	292	HTPX_BUCAI	P57406 buchiera ap
16	119.5	5.1	2294	YCF2_ARATH	P56766 arabidopsis
17	119	5.1	297	HTPX_STRGC	O30795 streptococc
18	116	5.0	293	HTPX_ECOLI	P23894 escherichia
19	113	4.8	284	HTPX_MERJA	O59076 methanococc
20	113	4.8	298	HTPX_BACSU	O31657 bacillus su
21	109	4.7	506	YCX2_EUDGR	P31921 euglena gra
22	109	4.7	595	YC86_MERJA	O58682 methanococc
23	107.5	4.6	1032	MT18_YEAST	P40469 saccharomyc
24	106.5	4.5	646	MTPL_HAEPV	O37319 heliothis a
25	106.5	4.5	1029	YFC5_YEAST	P43571 saccharomyc
26	105.5	4.5	873	COX1_ACACA	Q37370 acanthamoeb
27	105	4.5	1427	SRB8_YEAST	P25648 saccharomyc
28	104.5	4.5	585	YHDB_YEAST	P58732 saccharomyc
29	104.5	4.5	1100	LOS1_YEAST	P33478 saccharomyc
30	104	4.4	302	HTPX_AQUAF	O67798 aquilex aeo
31	104	4.4	530	CP14_CHICK	P79760 gallus gall
32	103.5	4.4	185	GU01_RAT	P35894 rattus norv
33	103.5	4.4	450	VATH_SCHPO	O14265 schizosacch

ALIGNMENTS

RESULT	1	ST24_YEAST	STANDARD:	PRT:	453 AA.
AC	P47154:	ST24_YEAST			
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	CAAX PRENYL PROTEIN 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC				
DE	ENDOPROTEASE 1) (PSEP 1) (A-FACTOR CONVERTING ENZYME).				
GN	STE24 OR APC1 OR YTR117W OR J2032.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Rose M., Koeltter P., Entian K.D.;				
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A. AND FUNCTION.				
RX	MEDLINE=97167681; PubMed=9015299;				
RA	Fujimura-Kamada K., Nouvet F.J., Michaelis S.;				
RT	"A novel membrane-associated metalloprotease, Ste24p, is required for				
RT	the first step of NH2-terminal processing of the yeast a-factor				
RT	precursor.";				
RL	J. Cell Biol. 136:271-285(1997).				
RN	[3]				
RP	FUNCTION.				
RX	MEDLINE=97218305; PubMed=9065405;				
RA	Boyarchuk V.L., Ashby M.N., Rine J.;				
RT	"Modulation of Ras and a-factor function by carboxyl-terminal				
RT	proteolysis.";				
RL	Science 275:1796-1800(1997).				
RN	[4]				
RP	FUNCTION.				
RX	MEDLINE=98365461; PubMed=9700155;				
RA	Tam A., Nouvet F.J., Fujimura-Kamada K., Slunt H., Sisodia S.S.,				
RA	Michaelis S.;				
RT	"Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal				
RT	proteolysis and COOH-terminal CAAX processing.";				
RL	J. Cell Biol. 142:635-649(1998).				
RN	[5]				
RP	FUNCTION.				
RX	MEDLINE=98393572; PubMed=9725832;				
RA	Boyarchuk V.L., Rine J.;				
RT	"Roles of prenyl protein proteases in maturation of Saccharomyces				
RT	cerevisiae a-factor.";				
RL	Genetics 150:93-101(1998).				
RN	[6]				
RP	FUNCTION.				
RX	MEDLINE=20158928; PubMed=10692417;				
RA	Schmidt W.K., Tam A., Michaelis S.;				
RT	"reconstitution of the Ste24p-dependent N-terminal proteolytic step in				
RT	yeast a-factor biogenesis.";				
RL	J. Biol. Chem. 275:6227-6233(2000).				
RN	[7]				

34	103	4.4	645	1	VP74_NPYAC	P15663 autographa
35	103	4.4	1101	1	KOE5_YEAST	Q08217 saccharomyc
36	102	4.4	577	1	TH71_SCHPO	O09684 schizosacch
37	102	4.4	2493	1	YBA4_YEAST	P35194 saccharomyc
38	100.5	4.3	260	1	YMI6_RECAM	O21266 reclinomona
39	100.5	4.3	1040	1	RIK1_SCHPO	Q10426 schizosacch
40	100	4.3	1325	1	MRP4_HUMAN	O15439 homo sapien
41	99.5	4.2	309	1	O7AA_HUMAN	O76100 homo sapien
42	99	4.2	485	1	YCL1_KLEPN	O48457 klebsiella
43	98.5	4.2	1024	1	PPOB_PLAFA	P21421 plasmodium
44	98.5	4.2	313	1	NW2M_RHISA	O99817 thripcephal
45	98.5	4.2	417	1	O65A_DROME	P82982 drosophila

RE CHARACTERIZATION. PubMed-10825201;
 RA MEDLINE-20285442; Boyartchuk V.L., Pictologlou E.A., Rozema D.,
 RA Treblewood C.E., Rine J.,
 RA Poulter C.D., Rine J.,
 RT "The Caax proteases, Atc1p and Rce1p, have overlapping but distinct
 RT substrate specificities." Mol. Cell. Biol. 20:4381-4392(2000).
 RL [1]
 RN SUBCELLULAR LOCATION.
 RP MEDLINE-98409630; PubMed-9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kameda K., Michaelis S.,
 RT "Endoplasmic reticulum membrane localization of Rce1p and Ste24p,
 RT yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal A-factor cleavage." J.
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR MATING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z49617; CAAB9647.1; -
 DR EMBL: U77137; AAB38271.1; -
 DR MEROPS: M48.001; -
 DR SCD: S0003878; STE24.
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolyase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum; Pheromone response.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT METAL 297 297 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 298 298 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 301 301 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 390 390 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 394 394 PROTON DONOR (BY SIMILARITY).
 SO SEQUENCE 453 AA; 52324 MW; 331C9AE2D7C99DA CRC64;

Query Match 99.7%; Score 2336; DB 1; Length 453;
 Best Local Similarity 99.8%; Pred. No. 3.7e-160;
 Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MFDTKTLIDHNPIMKLIISGFSIAOFSESYLYTYROYOKLSETRKLPVLEDEIDDEFTH 60
 DB 1 MEDKTLIDHNPIMKLIISGFSIAOFSESYLYTYROYOKLSETRKLPVLEDEIDDEFTH 60
 OY 61 KSRNYSRAKAFSTFGDYNNLAQKLVFKYDLPFKIMHMAVSLNAVLPVRFHNVSTVAQ 120
 DB 61 KSRNYSRAKAFSTFGDYNNLAQKLVFKYDLPFKIMHMAVSLNAVLPVRFHNVSTVAQ 120
 OY 121 SLCEFLGSLISLTVLVDLPVLSYSHVLEERKGFNKLTVQTLTMMKISLTLAVYAGGIL 180
 DB 121 SLCEFLGSLISLTVLVDLPVLSYSHVLEERKGFNKLTVQTLTMMKISLTLAVYAGGIL 180

OY 181 YLFKIFDKFPFDLMYIMVLFVVOIAMTIIPIFMFPMFNKFTPLEDGELKKSIESLA 240
 DB 181 YLFKIFDKFPFDLMYIMVLFVVOIAMTIIPIFMFPMFNKFTPLEDGELKKSIESLA 240
 OY 241 DRVGFPDLKIVIDGSKSSSNMAYFTGLPPTSKRIYLFPLVNSNSTDEITAVLAHEIG 300
 DB 241 DRVGFPDLKIVIDGSKSSSNMAYFTGLPPTSKRIYLFPLVNSNSTDEITAVLAHEIG 300
 OY 301 HMOKNHIYNAVIFSOHLFLFSLFTSYRMTSPNFGFLEKSTGSFVDPVITKPERI 360
 DB 301 HMOKNHIYNAVIFSOHLFLFSLFTSYRMTSPNFGFLEKSTGSFVDPVITKPERI 360
 OY 361 IIGFNLFDLTLPLECANQFVMSLISRTHEYOADAVAKKLGKONLCRALDLOIKNLST 420
 DB 361 IIGFNLFDLTLPLECANQFVMSLISRTHEYOADAVAKKLGKONLCRALDLOIKNLST 420
 OY 421 MNVDPLVSSYHSHPTLAERSTALDYVSEKKN 453
 DB 421 MNVDPLVSSYHSHPTLAERSTALDYVSEKKN 453
 RESULT 2
 ST24_SCHPO STANDARD: PRT; 474 AA.
 ID ST24_SCHPO
 AC Q10071;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-
 DE SPECIFIC ENDOPROTEASE 1) (PPEP 1).
 GN SPAC3H1.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Gentiles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED PROTEINS (BY SIMILARITY).
 CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z68144; CA92258.1; -
 DR MEROPS: M48.001; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 230 250 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT METAL 332 332 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 333 333 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).


```

Db 141 WSLYNTFVIEKRGHNOQTGFGMDAIKKFVVTQCIILPVSSLLIYIKIGDYFIYA 200
Oy 199 MPELVQVLAITIIIPVFMPMFNKFTPELDEGLKSTESLADRCFPLDKIFVIDGSKR 258
Db 201 WPELVVLAIVLTITADYADLPEDKFTPELDEGLKSTESLADRCFPLDKIFVIDGSKR 260
Oy 259 SSSHNAFTGLPFTSKRIYLFDTLVNSNS----- 287
Db 261 SSSHNAFTGLPFTSKRIYLFDTLVNSNS----- 319
Oy 288 -----TDEITAVLAHEIGHQWOKNHNIVNVIFSQLHTFLIFSLFTSIYNTSEFYNTGFF 341
Db 320 KKGCKNEEVLAVLGHGELHGMKLGHTVKNIIISQMSFICFELFVALLIGRKELEFAFGFY 379
Oy 342 LKSTGSPFDPVITKEFPIIIGFM-LFNDLLTPLECAMQFVMSLSRTHEYQADAVAKKL 400
Db 380 DSO-----PFLIGLIIIFQIFSPYNEVLSFCLTVLSRFEQADAFAKKL 425
Oy 401 GYKONICRALIDLOIKNLSTMNVDPLYSYHYSHPTLAERSTAL 444
Db 426 GKAKDLYSALIKLNKNDLGFVSDMLFMSMHSHPLERLQAL 469

RESULT 4
YHFN_BACSU
ID PA0769: 007612: STANDARD: PRT: 426 AA.
AC 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE METALLOPROTEASE YHFN (EC 3.4.24.-) (Psp23).
GN YHFN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (Jun-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-49 FROM N.A.
RX MEDLINE=89108019; PubMed=3145906;
RT Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the
RL Bacillus subtilis chromosome.";
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Y14083; CAA74535.1;
CC EMBL: M22914; AAB22830.1;
CC EMBL: 299109; CAB12869.1;
CC Subtilisin: BG11029; yHFN.
CC InterPro: IPR001915; Peptidase_M48.
CC InterPro: IPR000130; Zn_MpPeptidse.
CC Pfam: PF01435; Peptidase_M48; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hypothetical protein: Hydrolase; Metalloprotease; Zinc;
CC Complete proteome.
CC METAL 277 277 ZINC (CATALYTIC) (POTENTIAL).
CC ACT_SITE 278 278 POTENTIAL.
CC METAL 281 281 ZINC (CATALYTIC) (POTENTIAL).

```

```

FT CONFLICT 43 49 ASFMKPS -> LESTAQ (IN REF. 2).
SQ SEQUENCE 426 AA; 48874 MW; 0F69E7EFAA80143 CRC64.

Query Match 14.0%; Score 328; DB 1; Length 426;
Best local similarity 26.9%; Pred. No. 1.7e-16;
Matches 108; Conservative 76; Mismatches 152; Indels 66; Gaps 14;

Oy 81 LAQKLVFIKYDF-----PKIMMAVSL-----NAVLPVRHNMSTVAQSICF 124
Db 53 VAEQYSNVKNFLFELGVPLDMFLFVLLVSGVSKRIKKIKWEAVPFR-----LQTVGF 106
Oy 125 LGLLSLSLTVLDPLPSYSHFELEKFGFNKLTVOYLTMDIKSTLAVYAG--PLIY 181
Db 107 VPELSLITLTVLPEDWIGYQVSLD-VYINSTOTSMKADQVYISWISFPIFLCVLFY 165
Oy 162 LFLKIDKFPPELWYINW---FLFVQVLLAMTIIIPVIMPMFNKFTPELDEGLKSTIE 237
Db 166 WLKRHEKKMMLYAWLTVPEFLFLEFIQ-----PVILDPYNDPFLKKELESKIL 218
Oy 238 SIADRVGPRPLKIFVIDGSKRSHSNAYFTGLPFTSKRIYLFDTLVNSNSTDEITAVLAH 297
Db 219 ELADANIPADHVYEVNNSKTNALNAVYTGIG-ANKRIYVMDTTLNLDSEILFIIGH 277
Oy 298 EIGHQWOKNHNIVNVIFSQLHTFLIFSL-----FTSIYNTSEFYNTGFFLEKSTGSPVDP 352
Db 278 EMGHVYMKH-----YIGLAGLLVSLAGFYIDKLYRTVRLTSMHLEGRHDLALP 332
Oy 333 VITKEFPIIIGFMFNNDLTPLECAMQFVMSLSRTHEYQADAVAKKL-GYKONICRALI 411
Db 333 L-----LFLFVSLSFVAVTPFSNA-----VSRQYENKADQYIELTENEAIVTFQ 379
Oy 412 DQIKNLSTMNVDPLYSYHYSHPTLAERSTALDVSSEKKN 453
Db 380 DLAVGLSQVDPVPLVAKIFRGSHPSIMERIGH---AKREEN 417

RESULT 5
HHPX_PASMU
ID HHPX_PASMU STANDARD: PRT: 286 AA.
AC P57846;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HHPX HOMOLOG (EC 3.4.24.-).
GN HHPX OR PM0468.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE006083; AAK02552.1;
CC InterPro: IPR001915; Peptidase_M48.
CC InterPro: IPR000130; Zn_MpPeptidse.
CC Pfam: PF01435; Peptidase_M48; 1.

```

DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 KW Transmembrane: Hydrolase; Metalloprotease; Zinc; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT METAL 139 139 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACID SITE 140 140 BY SIMILARITY.
 FT METAL 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 286 AA; 31344 MW; 58FAA7841C300BCF CRC64;
 Query Match 6.9%; Score 161; DB 1; Length 286;
 Best Local Similarity 24.8%; Pred. No. 9.4e-05;
 Matches 76; Conservative 53; Mismatches 128; Indels 50; Gaps 13;
 OY 160 LMTMTKSLPLAVAGGILYLFKIDKPTDFLWYIMVFLVYQILMTIIPVIMP 219
 DB 15 LVFVNIILSLTGIOADATGLMLALFGSGS-----LISLFSLKTMALRAVGAEVIK 68
 OY 220 MENKFTPLEDGE--LKSSIESLADRVGFPDLKIFVIDGSKSSHSNAVETGIPFSKRIV 277
 DB 69 Q-----PRNDEKRLVNTVRSOAEKANI.PMDVAIY----HSEDVNAEATG-PSKNSLV 118
 OY 278 LFDT-LVNSNSTDEITAVLAHEIGHQKNHIVMYIF-SQHTFLIFSLFTSIYNTSEY 335
 DB 119 AVSTGLLRAMQDEAEVLAHEVAHAIKNGDMVTMLLGVLNTFVIF--VSRMIKKVSS 176
 OY 336 NTEGFFLEKSTGSEVDVYITKEPFIILIGFMLENDLLPLECAMQFVMSLISRTHEYQDA 395
 DB 177 NNDG---ESSGIYF--LVSWLEILFGFL-----ASMIAMFESRYREFRDA 219
 OY 396 YAKKGYGNQNCRALIDLOIKN-----ISTMNVDPL---YSSYHSHPLARSTAL 444
 DB 220 GSAKLVGHKMAIALQRLQRLHEPOELGQLAAFAINRGGLALFLMSHPLEKRIAL 279
 OY 445 DYVSEKK 451
 DB 280 QQLDSFK 286
 RESULT 6
 HTPX_HELPJ STANDARD; PRT; 310 AA.
 ID HTPX_HELPJ
 AC 09KRS4:
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
 GN HTPX OR JHP0861.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP MEDLINE FROM N.A. PubMed-9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO01516; AAO06444.1;
 DR InterPro: IPR001915; Zn_mtpeplase_M48.
 DR InterPro: IPR000130; zn_mtpeplase.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 KW Transmembrane: Hydrolase; Metalloprotease; Zinc; Complete proteome.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 55 75 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT METAL 166 166 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 167 167 BY SIMILARITY.
 FT METAL 170 170 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 310 AA; 35470 MW; 4188493E1D3C6395 CRC64;
 Query Match 6.4%; Score 150.5; DB 1; Length 310;
 Best Local Similarity 23.3%; Pred. No. 0.00058;
 Matches 67; Conservative 57; Mismatches 106; Indels 57; Gaps 11;
 OY 184 LKIDKPTDFLWYIMVFLVVO--ILAMTIPV--TIPMKNKFTPLEDGLKKSISLA 240
 DB 50 LMTQIIFPT-----ITYMFVAVIILVLCIONSSIMLSDEKLDPSKVSKEKQI 104
 OY 241 DRVGFPL-----DKIFVIDGSKSSHSNAVFTGIPFSKRIVLFDTLVNSSTDEI 291
 DB 105 HRLLELLEAKLHREPKLYIIN---APYNNAFASGDESNILALTSALIERLDREL 160
 OY 292 TAVLAHEIGHQKNHIVMYIFSQHTFLIFSLFTSIYRNTSPNTRCFPL--EKSGSF 349
 DB 161 KAVIAHELSHRIHNDIRLTCVGLISNIMLVANFSV-----FFGNKRNKSGAN 210
 OY 350 VDPVITKEPFIILIGFMLENDLLPLECAMQFVMSLISRTHEYQDAVAKKGY-KONLCR 408
 DB 211 LARMLWVLOIILPFLT-----LLQWYLSITREYMDSCAFLMHDKPKAIR 258
 OY 409 ALIDL-----QIKNLSTMNVDPLYSSYHS-HPTLAERSTAL 444
 DB 259 ALQKISNDYTNNDYKEIDKNSTRSAAYLFNEMESTHRSIKNRISL 305
 RESULT 7
 HTPX_ARCFU STANDARD; PRT; 335 AA.
 ID HTPX_ARCFU
 AC 030004;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
 GN HTPX OR AF0235.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP MEDLINE FROM N.A. DSM 4304 / ATCC 49558;
 RA SPRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Keine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon *Archaeoglobus fulgidus* .
 RL Nature 390:364-370(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001090; AAB90998.1; .
 DR MEROPS: M48.013; .
 DR TIGR: AF0235; .
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
 KW Transmembrane; Hydrolase; Metalloprotease; zinc; Complete proteome.
 FT TRANSMEM 9 29 POTENTIAL.
 FT TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 64 84 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 213 233 POTENTIAL.
 FT METAL 168 168 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 169 169 BY SIMILARITY.
 FT METAL 172 172 ZINC (CATALYTIC) (BY SIMILARITY).
 FT SEQUENCE 335 AA; 37855 MW; 94103E840A2AC68B CRC64;
 SQ
 Query Match 6.4%; Score 150; DB 1; Length 335;
 Best Local Similarity 21.8%; Pred. NO. 0.00069;
 Matches 71; Conservative 51; Mismatches 113; Indels 90; Gaps 13;
 QY 163 TDMIKSLTAVNIGPILYFLK---IFDKPFDLWYIMFLVVOILAMTIIPVIMP 219
 DB 44 TSMVLACMILAIISAIIITLILAVGVISF-----YGLIIFLLINLIMLYLSTYIINL 98
 QY 220 MFNKTPLDEGLKKSISLADRVGFPDLKIFVIDGSKRSSNAYFTGLPETSRIYLF 279
 DB 99 SYG---AORDERLQWVSVARLNVKPKKAVV---RSPNNAFVAGNFTGKFAVVS 150
 QY 280 DTVVNSNSTDEFTAVLAHEIGHMOKNHTVNVNIFSQLHTPLIFSLFTSYNRTSYNFG 339
 DB 151 ESLMMLSQBELAVIGHIGH--HKRDNAVY-----LLEGLPSVI---FY--LG 195
 QY 340 FLEKST-----GSFVDPVITKEPPIIGMLFNDLTPLECAMQFWSLISRTHEYOA 393
 DB 196 YALLHSSMRDRRGAQLAIGIA-AVIVSFI-----VQILVAFSRREYVA 241
 QY 394 DAYAKKLYKONLCALIDL-----QI 415
 DB 242 DFEGRATNKAMORSIAKIHLEFYHRYPDYLAIDSKRFLFIYAFNAVAEPIRADI 301
 QY 416 KNLSTMANVDPYSSYHSHPTLAER 440
 DB 302 EALKMKVSPIT-QEFLSTHPLPKR 325
 RESULT 8
 HTPX_HAEIN STANDARD: PRT: 283 AA.
 AC P44840: HTPX_HAEIN
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE PROTEASE HTPX (EC 3.4.24.-).
 GN HTPX OR HI0720.
 OS Haemophilus influenzae.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

CC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Flier L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32755; AAC22378.1; .
 DR MEROPS: M48.014; .
 DR TIGR: HI0720; .
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Transmembrane; Inner membrane; Hydrolase; Metalloprotease; zinc;
 KW Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 190 210 POTENTIAL.
 FT METAL 139 139 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 140 140 BY SIMILARITY.
 FT METAL 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
 FT SEQUENCE 283 AA; 30780 MW; FA062B1BE0E0818 CRC64;
 SQ
 Query Match 6.3%; Score 147.5; DB 1; Length 283;
 Best Local Similarity 26.5%; Pred. NO. 0.00086;
 Matches 73; Conservative 42; Mismatches 101; Indels 59; Gaps 12;
 QY 198 IMVLEFVQILAMTIPFIMPENKFTPLE--DGE-----LKSISLADRV 243
 DB 37 IMALLFG---FAGSLISFL---SKTVALRSVGEVITQPRQTERLIDTVSRQAKA 89
 QY 244 GFLDRTVIDGSKRSSNAYFTGLPETSRIYLFDTLVNSNSTDEFTAVLAHEIGHQ 303
 DB 90 GIPMDVAIV---HSPDVNAFATGATKNSLVAVSGLLNNMTAEAEVAIAHEISHIS 145
 QY 304 KNLIVNVYIF-SQLHTPLIF-----SLTSLYRNTSFNFGFLEKSTGSFVDPVITKE 357
 DB 146 NGDMVYKALLOGVLTATFVIFLSRYIATRAVASSRNNGEET-----RSQGIPLVSVYLE 199
 QY 358 FPIIGFMLFNDLTPLECAMQFWSLISRTHEYQADAYAKKLGKYLKONLCALIDLQ--- 414
 DB 200 -----MLREVLASIT--AMWF-----SRKRRADAGSASLVGKEMMALORLOOLH 245
 QY 415 -----KNLSTMANVDPYSSYHSHPTLAERSTAL 444
 DB 246 EPQNLGSLNFMINGKRSLEFMSHPLEKRIEAL 280

```
RESULT 9
HTPX_HELPY
ID HTPX_HELPY STANDARD: PRT: 310 AA.
AC 025582:
DT 15-JUL-1999 (Rel. 38, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR HP0927.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McEnney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL NCBI MEDLINE=93310339; PubMed=10382966;
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```
OY 293 AVLAHEIGHMOKNHNIVNVVIESOLHTFLIFLSTIYRNTSFYNTFGFEL--EKSTGSFV 350
DB 162 AVIAHELHSHKNDIRLTMWCGILSNIMLVANSFY-----FFMGKRNKSGANL 211
OY 351 DPVITKEPPIITIGFMLFNDLLTPLECAMQFVMSLISRTHEYOADAYAKKIGY-KONLCRA 409
DB 212 ARMILWQLIILPFLTL-----LLQWLSRFRREYMAVDSGAFLHDKKPMIRA 259
OY 410 LIDL-----QIKNLSTMNVDPLSYTHYS-HPLAERSTAL 444
DB 260 LQKISNDYTNNDYKEIDKNSTRSAAYLFNMAEFTHPSIKNRIOSL 305

RESULT 10
HTPX_AERPE
ID HTPX_AERPE STANDARD: PRT: 366 AA.
AC Q9YD67;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR APE1045.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=93310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamizawa M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```


DR EMBL: AL445065: CAC11990.1: -
DR InterPro: IPR001915: Peptidase_M48.
DR InterPro: IPR00130: Zn_M1peptidase.
DR Pfam: PF01435: Peptidase_M48; 1.
DR Prosite: PS00142: ZINC_PROTEASE; FALSE_NEG.
KM Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT METAL 146 146 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 147 147 BY SIMILARITY.
FT METAL 150 150 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 317/AA; 35616 MW; 535EBATF733CDEE CMC64;
Query Match 5.6%; Score 130.5; DB 1; Length 317;
Best Local Similarity 17.5%; Pred. No. 0.016;
Matches 60; Conservative 68; Mismatches 128; Indels 87; Gaps 12;
QY 156 LTVQDMITDMKSLTLAVATGCPILYLFKIFDKFPDPLKIMFLVVOILANTIPV 215
Db 4 MSIRKLSSLMLGIAL-FAIGGIYA-----ILVYIGNIAIALL-NVLLPI 50
QY 216 FIM-----PMNKPTPLEDGLKKSIESLADRVGFPIDK 249
Db 51 FIIINDIYQWLEGPIMIGAVRTHKIQPEPEXNMLTST-----VSEVARNNNIRPE 102
QY 250 IEVIDGSKRSSHNAAYFTGLPTFSKRIVLEDTLVNSSTDEITAVLAHEIGHMOKNIIVN 309
Db 103 VYIAE---VDMPAFAAGSPAGRIARIVTGLLRILPEDEIKAVIGHEMGLHRIDVEM 158
QY 310 MVIISOHTELIFSLFISIVRNTSYNFFGFLEKSTGSFVDVPTKFPPIIGMFL-- 367
Db 159 LLAIGLIPTLFFGYTILFSGERNGGIYLL-----AIIAAMASFLEFRLIAF 210
QY 368 -----NDLIT-----PLECAMOFVSSLISRTHEY-----QADYAKKLGKYQ 404
Db 211 NRMESYADVNSALTVDGAEKIQTALAKIVAATGRTGRYRRRRNPSSVTEMLFSPSN 270
QY 405 NLCRALIDLQ--IKNLSTMNVDPLYSSHYSHPTLAEKSTALD 445
Db 271 PNQSANEDYRKLLEEMKTAKVS-LEADFSDHPHPAKRIQRLE 312
RESULT 14
YCF2_MARPO STANDARD; PRT; 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL 259 KDA PROTEIN YCF2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiales; Marchantiaceae; Marchantia.
OX NCBI_Taxid=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyama K.;
RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.
NN [2]
RP COMPLETE GENOME.
RA Ohyama K., Fukunawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT liverwort Marchantia polymorpha chloroplast DNA";
RT Nature 322:572-574(1986).
CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X04465; CAA28078.1; -
 DR PIR: A05037; A05037.
 DR PIR: S01591; S01591.
 DR Mendel: 5295; Marko.ycf2.1.
 DR InterPro: IPR001939; AAA_subfam.
 DR Pfam: PF00004; AAA.1.
 KM Chloroplast: Hypothetical protein.
 SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 5.4%; Score 127.5; DB 1; Length 2136;
 Best Local Similarity 21.3%; Pred. No. 0.26;
 Matches 113; Conservative 73; Mismatches 171; Indels 173; Gaps 28;

OY 1 MFDLTIIDHPNIP-----WKLIIT-----SGFSIAQSFSESYLYTKQYOKLSETKLPVYLE 51
 DB 156 LIDRSILKKKEIILHMKFLVLEQIOSNMKISE-----ESLSELEKI--VLE 201
 OY 52 DEIDDEPH-----KSNRYSAKAKKFSFGVY 79
 DB 202 OKNIDELKHFEFYINOKIYNNMWEYFYSIFINQKIDIKSNKYNSTSGFEF---L 258
 OY 80 NLAOKLVFIKYDEPKIMHMAVSLNANVLPVRFHNVSTVAOSLCFLGLSLSLTVDLPL 139
 DB 259 AFCEKLEF-----EVEFLSKPNNNNLQMKLNCLENFSLDLF-----CLNNKL 302
 OY 140 SYTHFVLEEFKGNKLTQVLMITDMIKSLTFLAIAIGPILYLFKLPDKFPDPLWYIM 199
 DB 303 PWNKKIKFKNLQNFNENSKL-----IESFLLKIKGN---LYFKNIIEFVT--W--- 347
 OY 200 VFLVVOIAMTIIPIVFMPEKFTPLEDEGLKKSIESLADRGFELDKIFVIDGS--- 256
 DB 348 -----QSYKKDCID-----FNKFNELNSEIYIKIEELSDYIKFSKILVEGKSK 395
 OY 257 ---KRSSHNAFTGLP--FTSKRIVLEDTLVNSNSTDEITAVLAHEIGHW-QKNHIVNM 310
 DB 396 TIIQSFNNNIYKKKLSINFNFIYED-----SNNL-----LFDLKKNIYINN 441
 OY 311 VIFSOLHTFLSLFTSYRTSYNTF-GFFLEKSTGSGF-----VDPVITKE 357
 DB 442 KPF--LKSFLIYS-----SISNOFILFFKQNSKSFKNLKKNSKDVITNVEFSK 490
 OY 358 PFIIG-----FMLFNDLPLPLCAMQFVMSLSRTH---EYQADAVAKKLGKONICR 408
 DB 491 NKIEINFSKSIYAFELISINELDKNFYINKISLKNKKOKRPFYLNKIKSSDNF-- 548
 OY 409 ALIDL-OIKNLSTMNVDPLSSYHYSHPTLAERSTALDYVE---KKK 453
 DB 549 RFIULMKIKN-----YSSQOFVSNNSFLNPAPEILQONYYLKKKN 589

RESULT 15
 HTPX_BUCAL STANDARD; PRT; 292 AA.
 ID HTPX_BUCAL
 AC P57406;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
 GN HTPX OR BU321.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OC symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Matanabe H., Hattori M., Sasaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RL Buchnera sp. Aps. ";
 RL Nature 407:81-86(2000).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AP001119; BAB13029.1; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR00130; Zn_Mtpepdse.
 DR Pfam: PF01435; Peptidase_M48.1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
 FT TRANSMEM 4 24
 FT TRANSMEM 32 52 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 193 213 POTENTIAL.
 FT METAL 139 139 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 140 140 BY SIMILARITY.
 FT METAL 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 292 AA; 32762 MW; DA0135106794BA30 CRC64;

Query Match 5.1%; Score 120; DB 1; Length 292;
 Best Local Similarity 23.1%; Pred. No. 0.083;
 Matches 68; Conservative 48; Mismatches 101; Indels 78; Gaps 14;

OY 200 VFLVVOIAMTIIPIVFMPEKFTPLEDEGLKKSIESLADRGFELDKIFVIDGS--- 256
 DB 20 LILFLTGIOSTWITVIGLIMSGLFGSGSILSLIKMIALSVNGEITTHRNEVESWLI 79
 OY 241 DRV-----GEPIDKIFVIDGSKRSSHNAFTGLPFTSKRIYLFIDLVSNSDETA 293
 DB 80 NTVRQOSTQKGIIMQIAVY---HADDINAFATGARNASALIAVSTGLLNMTHHEEA 135
 OY 294 VLAHEIGHMOKNHIYVM-VIFSOLHTFLIF-SLFTSIY-----RMTSEYNTGFG 340
 DB 136 VIAHEISHIANGDMITMTLVGVVNTFVIFISRLSOLISVMSNNRRENTTEKNSFY 195
 OY 341 FLEKSTGFSVDPVITKEPPIIGFMLFNDLPLPLCAMQFVMSLSRTHETQADAVAKL 400
 DB 196 FL---VSTFLE-----LIFGIL-----ASITMWFSSRRREFYADASSAKM 232
 OY 401 GYKONLCALDLOIKNLSTMNVDPL-----YSSYHYSHPTLAERSTAL 444
 DB 233 VGRKMTAALNRLKTSR-EPOSDSMIAFCLNGKSKSLFLKPSHPSLNRIEAL 286

Search completed: February 12, 2002, 09:32:24
 Job time: 340 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:25:34 ; Search time 26.02 Seconds

(without alignments)
1326.175 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342

Sequence: 1 MFDLKTIIDHPNIPWKLIIIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BLOSUM62

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database: PIR-68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2336	99.7	453	2	S57140	probable membrane
2	903	38.6	474	2	T38737	probable CAAX pren
3	819.5	35.0	459	2	C85017	probable CAAX pren
4	654	27.9	316	2	T01712	hypothetical prote
5	612.5	26.2	442	2	T18917	hypothetical prote
6	501.5	21.4	395	2	F81343	probable integral
7	459	19.6	407	2	D71862	probable zinc meta
8	454	19.4	407	2	F64567	zinc metalloprotei
9	378	16.1	428	2	E84192	zinc metalloprotei
10	328	14.0	426	2	C69831	conserved hypotet
11	197	8.4	338	2	H75127	CAAX prenyl protei
12	152	6.5	291	2	D83293	heat shock protein
13	150.5	6.4	310	2	D71878	stress protein - H
14	150	6.4	335	2	C69279	heat shock protein
15	148.5	6.3	279	2	E81951	probable membrane
16	148.5	6.3	347	2	E81154	heat shock protein
17	147.5	6.3	283	1	H64088	heat shock protein
18	142.5	6.1	326	1	G64635	heat shock protein
19	139	5.9	366	1	F72703	probable heat shoc
20	138.5	5.9	258	1	C69175	heat shock protein
21	136	5.8	287	2	E82239	heat shock protein
22	135.5	5.8	289	2	B82534	heat shock protein
23	130.5	5.6	292	2	C71070	probable heat shoc
24	129.5	5.5	304	2	F83987	heat-shock protein
25	127.5	5.4	2136	2	A05037	hypothetical prote
26	123	5.3	315	2	E69365	hypothetical prote
27	122	5.2	3119	2	T18414	protein g377 - mal
28	120.5	5.1	289	2	A75094	heat shock protein
29	120	5.1	292	2	E84967	heat shock protein

30	119	5.1	297	2	T48855	probable heat shoc
31	117.5	5.0	324	2	B86672	permease [imported
32	116	5.0	293	1	A43659	heat shock protein
33	116	5.0	293	2	G85794	heat shock protein
34	113	4.8	284	1	H64509	heat shock protein
35	113	4.8	298	1	G69862	heat-shock protein
36	110.5	4.7	355	2	T29251	hypothetical prote
37	109.5	4.7	312	2	G70114	phosphatidate cyti
38	109.5	4.7	669	2	S76474	hypothetical prote
39	109	4.7	506	2	S34457	hypothetical prote
40	109	4.7	595	2	E64460	hypothetical prote
41	107.5	4.6	1032	2	S53571	hypothetical prote
42	106.5	4.5	645	2	T41876	p74 orf138 - Bomby
43	106.5	4.5	1029	2	S56229	probable membrane
44	106	4.5	842	2	T04880	potassium transpor
45	105.5	4.5	478	2	T11318	NADH dehydrogenase

ALIGNMENTS

Swiss Prot P0471574

RESULT 1	
S57140	probable membrane protein YJRI17W - Yeast (Saccharomyces cerevisiae)
N:Alternate names: -hypothetical protein J2032	
C:Species: Saccharomyces cerevisiae	
C:Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 06-Feb-1998	
C:Accession: S57140	
R:Rose, M.; Koetter, P.; Entian, K.D.	
submitted to the Protein Sequence Database, September 1995	
A:Reference number: S56848	
A:Accession: S57140	
A:Molecule type: DNA	
A:Residues: 1-453 <ROS>	
A:Cross-references: EMBL:Z49617; NID:G1015836; PID:G1015837; MIPS:YJRI17W	
C:Genetics:	
A:Gene: SGD:STRE24	
A:Cross-references: SGD:S0003878; MIPS:YJRI17W	
A:Map position: 10R	
C:Keywords: transmembrane protein	
Query Match	99.7%; Score 2336; DB 2; Length 453;
Best Local Similarity	99.8%; Pred. No. 4.9e-171;
Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY 1 MFDLKTIIDHPNIPWKLIIISGFSIAQSFESYLYROYOKLSEKLPVLEDEIDETFH 60	
Db 1 MFDLKTIIDHPNIPWKLIIISGFSIAQSFESYLYROYOKLSEKLPVLEDEIDETFH 60	
QY 61 KSRNYSRAKAKFSIFGDVYNLAOKLVFIKYLPFKIMHMAVSLNAVLPVRFHNVSTVAQ 120	
Db 61 KSRNYSRAKAKFSIFGDVYNLAOKLVFIKYLPFKIMHMAVSLNAVLPVRFHNVSTVAQ 120	
QY 121 SLCEFLGLISLSTVVDLPVLSYSHVLEKSGPKNTVQOLITTMIKSLTAVYIAGPIL 180	
Db 121 SLCEFLGLISLSTVVDLPVLSYSHVLEKSGPKNTVQOLITTMIKSLTAVYIAGPIL 180	
QY 181 YLFLEKIDFKPTDFLWYIMVFLFVQILAMTIIPIVFIIPMKFNKFTPLEDGLKKSIESLA 240	
Db 181 YLFLEKIDFKPTDFLWYIMVFLFVQILAMTIIPIVFIIPMKFNKFTPLEDGLKKSIESLA 240	
QY 241 DRVGFPLDKIFVIDGSKRSHSNAYFGLPPTSRIYVFDPLVNSNSTDEITAVLAHEIG 300	
Db 241 DRVGFPLDKIFVIDGSKRSHSNAYFGLPPTSRIYVFDPLVNSNSTDEITAVLAHEIG 300	
QY 301 HMQNHNHYNVYFSQLHTEFLFSLFSTSYRNTSYNPGFLEKSTGSFVDPVITKEPPI 360	
Db 301 HMQNHNHYNVYFSQLHTEFLFSLFSTSYRNTSYNPGFLEKSTGSFVDPVITKEPPI 360	
QY 361 TIIGMLFNDLITPLECAMQFVMSLISRTHEYOQADAVAKKLGKYNKLRALIDLIQKNLST 420	
Db 361 TIIGMLFNDLITPLECAMQFVMSLISRTHEYOQADAVAKKLGKYNKLRALIDLIQKNLST 420	

OY 421 MNVDLYSSHYSHPTLAERSTALDYSEKKKN 453
 |||||||
 Db 421 MNVDLYSSHYSHPTLAERSTALDYSEKKKN 453

RESULT 2

T38737
 probable CAAX prenyl proteinase 1 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
 R:Accession: T38737
 R:Genetic: S. Church, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, December 1995
 A:Reference number: 221808

A:Accession: T38737
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-474 <GEN>
 A:Cross-references: EMBL:Z68144; PIDN:CAA92258.1; GSPDB:GN00066; SPDB:SPAC3H1.05
 A:Experimental source: strain 972h-; cosmid c3H1
 A:Gene: SPDB:SPAC3H1.05
 A:Map position: 1
 A:Introns: 35/3

Query Match 38.6%; Score 903; DB 2; Length 474;
 Best Local Similarity 40.4%; Pred. No. 1.7e-61;

Matches 184; Conservative 96; Mismatches 153; Indels 22; Gaps 6;

OY 4 LKTIIDHPNIMKLIISGSIAGSFESYLYROYOKLSEKRLPVLDEIDDEFHRSR 63
 |||||
 Db 37 LMHIDIGFPMKIVAGSISGKAYMDLYLRQVRYLLREKPAIILAHYDEKKYQAL 96
 OY 64 NYSRAKAKSIFGQVYNLAOKLVFKYDLPKIMHMA---VSLNAVLPVRFHNVSTVA 119
 |||||
 Db 97 SYARDKSWFSTIVSTFTLAVDLILIKYDLSYLMNITKFPMMDKL-AASSRFSLSITSIT 155
 OY 120 QSLCEFLGLSLTLVDLPISYSHFVLEKFGFKLTVOLMTDMKSLTALAIGPPI 179
 |||||
 Db 156 HSCVFMELTLESRLIQIPNLYSTFVLEKFGFKSLKLFVIDLDELSTGLMSV 215
 OY 180 LYLFLKIPDKPTDPLATVIFLFPVOILAMTIIIPVFMPEKPTPEDEGLKKSISL 239
 |||||
 Db 216 VGVFVKILTKGDNFIMKAGAYIVFGILDTIAPSLIMPTFYKFTPLENSLKTQIDEL 275
 OY 240 ADRVGFPLDKIFVIDGSRSSSHNAYFTGLPFTSKRIYLPDTLVNSSTDEITAVLAHEI 299
 |||||
 Db 276 AASINFPKRLKYVIDASRSTHNAFYGDPW-NKGIYLPDTLVKNHPEPELIALIGHEL 334
 OY 300 GHMOKNH-IVMMVIFSQLHTFLISLFTSTRNTSFYTFGFELEKSGSFVDPIYTKEF 358
 |||||
 Db 335 GHMWSHNLMTIIDYGSLEHLP-LFAAFIRNNSLYTSFNITEK----- 379
 OY 359 PIIGFMLENDLTPLECAMQFVMSLSRTHGYOADAAYAKKGYKONCRALIDLQIKNL 418
 |||||
 Db 380 PVIYGLLFSDALGPLSSILTFASKNYSRLCEYQADAPAKOLGAKDGLGIRIHNDNL 439
 OY 419 STMNVDPLYSSHYSHPTLAERSTALDYSEKKKN 453
 |||||
 Db 440 SPLEFDSLTSYVSHSHPLIVDLRLNADYTLTKNN 474

RESULT 3

C85017
 probable CAAX prenyl proteinase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

R:Accession: C85017
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: C85017
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-459 <STO>
 A:Cross-references: GB:NC_001268; NID:g7267629; PIDN:CAB80941.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: AT4G01320
 A:Map position: 4

Query Match 35.0%; Score 819.5; DB 2; Length 459;
 Best Local Similarity 38.1%; Pred. No. 4e-55;

Matches 184; Conservative 73; Mismatches 155; Indels 71; Gaps 9;

OY 13 IPWKIIISGSIAGSFESYLYROYOKLSEKRLPVLDEIDDEFHRSRYSRAKAF 72
 |||||
 Db 3 IPFMTVVGFMIVMVFETVLDRLQTLKLPTEKTLGVISQKFEKSRAYR----- 56
 OY 73 SIFGVNLAQKIVIK-----YDLPKIMHMAVSLNAVLP-VRFHNVSTV 118
 |||||
 Db 57 DITENFNICSYFHFVHEPVTIIMDSALFPGILPFMKMS---GAVLPRLGLDPENET 112
 OY 119 AQSICEFLGLSLTLVDLPISYSHFVLEKFGFKLTVOLMTDMKSLTALAIGPPI 178
 |||||
 Db 113 LHTLSFLAGVMTMSQITDLPESLSTFVIESRHGFKNTIMFIDMKGTFLSYLGLP 172
 OY 179 ILVFLKIPDKPTDPLATVIFLFPVOILAMTIIIPVFMPEKPTPEDEGLKKSIS 238
 |||||
 Db 173 IVAALIFIVOKGPLYALVIAFMFLLSLVMWITPVILAPLFNKFPLPGDDLEKTEK 232
 OY 239 LADRVGFPLDKIFVIDGSRSSSHNAYFTGLPFTSKRIYLPDTLVNS-NSTDEITAVLAH 297
 |||||
 Db 233 LASSKFPKRLKYVIDASRSTHNAFYGDPW-FKKRIYLPDTLVKNHPEPELIALIGHEL 291
 OY 298 EIGHMOKNHIVMMVIFSQLHTFLISLFTSTRNTSFYTFGFELEKSGSFVDPIYTKEF 357
 |||||
 Db 292 ELGHMOKNHIVMMVIFSQLHTFLISLFTSTRNTSFYTFGFELEKSGSFVDPIYTKEF 337
 OY 338 PIIGFMLENDLTPLECAMQFVMSLSRTHGYOADAAYAKKGYKONCRALIDLQIKNL 416
 |||||
 Db 338 -PVLIGLIIIFQHTVPIPLHVSFGLNLSRAFEFGADFAVAKLGRALPVLKLYQRE 396
 OY 417 -----NLSTMNVDPLYSSHYSHPTLAERSTALDYSE 449
 |||||
 Db 397 DNNRTQYVTSICVTHLNGFVGILOENLSAMNDPLYSAYHSHPLVERLAID--GE 454
 OY 450 KKK 452
 ||
 Db 455 DKK 457

RESULT 4

T01712
 hypothetical protein A.IG002N01.21 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999

R:Accession: T01712
 R:Schaefer, P.; Magg, L.
 submitted to the EMBL Data Library, June 1997
 A:Description: The sequence of A. thaliana IG002N01.
 A:Reference number: 214407

A:Accession: T01712
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-316 <SCH>
 A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191141
 C:Genetics:
 A:Experimental source: cultivar Columbia
 A:Map position: 4
 A:Introns: 11/1; 40/3; 62/3; 95/3; 133/3; 170/3; 191/3; 222/3; 246/3; 278/3
 A:Note: A_IG002N01.21

Query Match 27.9%; Score 654; DB 2; Length 316;
 Best Local Similarity 34.1%; Pred. No. 1.1e-42;
 Matches 154; Conservative 54; Mismatches 93; Indels 150; Gaps 9;

13 IPWKIIIGFSTIAQSFESYLTROYOKLSETKLPVLEDEIDETFKSRNYSRAKAF 72
 Db IPFMETVVGFMIVMYIFETYLDRQ---LFAIKLPYL----- 36

73 SIFGVYVLAQKIVFIRKIDLPFKIMHMAVSLNVLPRFHMVSTVAOSLCFLGLLSLS 132
 Db -----PK----- 39

133 TLVDLPISYSHFVLEKFGFNKLTVOIMTDMIKSLTAAVAGGPILYFLKIPDKPPT 192
 Db LITDLPESLTFTVIESHGKNTIMFIRMIGTFLSYLGGPIYAAITFIYQKGP 99

193 DFLWYVLEFVVOILMTIIPVIMPENKFTPLEDELKKSIESLADRVGFPLDKITFV 252
 Db YLAIYMAFMFILTSLVMTIYVPLIAPLNKFTPLPDGDLREKIKLASLKFLPKLFLV 159

253 IDGSRSSHNAFTGTFPTSKRIYLPDTLVNS-NSTDEITAVLAHEIGHMOKNIIVMV 311
 Db VDGSTRSSHNAVMYGF-FKKRKRIYLYDPTLIQOCKNEDEIYAVLAHEIGHMOKLNH----- 213

312 IFSOLHTFLIFSLFTSIYRNTSEYNTFGFLEKSTGSPVDPVITKEPFIIGFMFLNDL 371
 Db -----TTSYFAVQHT----- 225

372 TPLECAMQVMSLISRTHEYADAVAKKGYKONICRALIDLOK-----NLSTM 421
 Db IPLHLVSGFLNLSRAEFQADAFVAVKGLDRLPALKVQVREDNNRTOEENLSAM 285

422 NVDPYSSYHSHPTLAERSTALDYSEKK 452
 Db NTDPYSAHYSHPELVRLAID--GDDKK 314

286 NTDPYSAHYSHPELVRLAID--GDDKK 314

RESULT 5
 T18917
 hypothetical protein C04F12.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18917
 R:Lloyd, C.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19044
 A:Accession: T18917
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-442 <WIL>
 A:Cross-references: EMBL:Z81461; PIDN:CAB03839.1; GSPDB:GMD0019; CESP:C04F12.10
 A:Experimental source: clone C04F12
 A:Gene: CESP:C04F12.10
 A:Map position: 1
 A:Introns: 30/2; 77/3; 318/1; 340/3

Query Match 26.2%; Score 612.5; DB 2; Length 442;
 Best Local Similarity 32.1%; Pred. No. 2.5e-39;
 Matches 150; Conservative 82; Mismatches 155; Indels 81; Gaps 10;

25 AQSFESYLTROYOKLSETKLPVLEDEIDETFKSRNYSRAKAFSIFGDVYNLAQ 83
 Db ALFLMDQITTFROYKAKHNAVKRPVEKVELGEEEDYKKARYKIDNHLGFFHSWFNQL 75

84 KLVEIKYDLFPKIMHMAVSLNVLPRFHMVSTVAOSLCFLGLSSLTVDPLSYSS 143
 Db LTAQIGYGFYFWMTATSS-----YPLHVA-----VLSINSIETIIDLPMWLXS 121

144 HFVLEEKFGFKLTVOIMTDMIKSLTAAVAGGPILYFLKIPDKPPTDPLWYIMVLF 203
 Db ----- 11

Db 122 FFIIDAHGFNKOTIGFVVDKIKKMLVGFALTMPLIVYGIEMIYVNGPFYVIMLFVS 181

204 VQOIIAMTIIPVIMPENKFTPLEDELKKSIESLADRVGPPDKIPIYIDSKSSSHN 263
 Db VVLLIMTYIPTFPIAPLFDKYPPLDGLKTIKIDLAASLYPTELIVVNGSKRSASHN 241

264 AFTGLPFTSKRIYLPDTLVNS-NSTDEITAVLAHEIGHMOKNIIVMV 296
 Db AVMYGF-WKNRKRIYLYDPTLIQOCKNEDEIYAVLAHEIGHMOKLNH----- 300

297 HEIGHMOKNIIVMVISQTLPTFLSFTSIYRNTSEYNTFGFLEKSTGSPVDPVIT 355
 Db HELGHMALWHTLINLVI-TEVNLFPFSAFVGYFVKEALYQGFY----- 344

301 HELGHMALWHTLINLVI-TEVNLFPFSAFVGYFVKEALYQGFY----- 344

356 KEPIIIGFMFLNDLTPLECAMQVMSL-----SRTHEYQADAAKGLKGYK 404
 Db HDTPVIGMMLT-----FOFVLAALYNOLASIGVYHSRSAFQADERANLGHGE 394

405 NLGALIDLOKNIILSTMNVDPYSSYHSHPTLAERSTALDYSEKK 452
 Db NLGALTLGVNLSMPINDSLYSCYTHRPVVERVAARAFQAKNK 442

395 NLGALTLGVNLSMPINDSLYSCYTHRPVVERVAARAFQAKNK 442

RESULT 6
 F81343
 probable integral membrane zinc-metalloproteinase Cj0723c [imported] - Campylobacter
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: F81343
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
 C.W.; Quail, M.; Rajadipram, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; B
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: A81250; MUID:20150912
 A:Accession: F81343
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-395 <PAR>
 A:Cross-references: GB:A139076; GB:A111168; NID:g6968128; PIDN:CAB72997.1; PID:g69
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0723c

Query Match 21.4%; Score 501.5; DB 2; Length 395;
 Best Local Similarity 31.2%; Pred. No. 6.7e-31;
 Matches 135; Conservative 88; Mismatches 154; Indels 55; Gaps 12;

18 IISGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETFKSRNYSRAKAFSIFGD 77
 Db LLSWISYAOIRPLEKDKOAOILSEKDYQANADAIENE-----KELFSN 59

13 LLSWISYAOIRPLEKDKOAOILSEKDYQANADAIENE-----KELFSN 59

78 VYNLAQKLVFIKYDLPKIMHMAVSLNVLPRFHMVSTVAOSLCFLGLSSLTVDL 137
 Db FYNLISIAWIGGF---LYLKEILLN-----NTRENPLFLSLIINSILN 106

138 PLSYSHVLEKFGFNKLTVOIMTDMIKSLTAAVAGGPILYFLKIPDKPPTDPLMY 197
 Db PLSYSEFIKDKAHGFSNMVTKFLIKDPMKSLTLILGFLLYLTCYDFEFG--FMW 164

198 IMVFVVOILAM--TIIPVIMPENKFTPLEDELKKSIESLADRVGPPDKIPIYIDG 255
 Db IAFIFACIITVILNLYPTLAPLFFNKKKIKDDELNKLKKSILKKGCGFSANGVYVDA 224

256 SKRSSHNAFTGTFPTSKRIYLPDTLVNSNSTDEITAVLAHEIGHMOKNIIVNISQ 315
 Db SKRDKRLNAYFGCL-FKSKRVVLPFTLLKALNERELLAVLGHGHPVHKDILKALFNGA 283

316 LHTFLISLFTSIYRNTSEYNTFGFLEKSTGSPVDPVITKEPFIIGFMFL--FNOLLT 372
 Db ITMFLFFVFANL---PEF---VYLE---SHLEGVNGVAFALL--FILANIFSPLIS 329

373 PLECAMQVMSLISRTHEYADAVAKKGYKONICRALIDLOKNIILSTMNVDPYSSYH 432

OY 100 AVSLNLAIVLPREHNVSTVAQSLCE-----GLLSLSTLYD-----LPLSYSH 144
DB 68 SVYLAGAVLGLV--TGAFGAANAFAVDVPGSDVLGAVALLGATVMAQVAPFEDAVET 125
OY 145 FVLEERKGFNKLTVQWITMIMIKSLTIA-----VAIGPILYLFLEKIDKPEPTDELAMYIM- 199
DB 126 FGVSASAGFNDQSPALFARQALLSAGLAVFVAVLGAVALVAALAE-----WNVVAA 179
OY 200 -----VLFVAVQILAMTIIPEVIMPENKFTPLEDEGLKKSIESLADRVGFPDLKIFVID 254
DB 180 TGIAGVFLATQVL-----VRRVAVMPLFEDPDVDEGLRDAIEDVDFRAGFACQVYVMN 235
OY 255 GSKRSSHNAVFTGLPFTSRIVLFDLVNSSTDEITAVLAHIGMOKNHIYNNMIFIS 314
DB 236 ASSRSQSNMFFTEFGAT-KRVVLEFDLIDMGETELQAVLAHLMKNNIQTGAA 294
OY 315 QLHTEFLFSLFTSYRTSYNTSPNTEGFLKSTGSPVDPVITKEPPIIGFLFNDLITPL 374
DB 295 TLGAGVLFVASRLDAGWLYGMG-----VPEQP-AAAGLLAGVWLQPL 338
OY 375 ECAMQFVMSLISRTHEQADAYA-KKLGKONLCRALIDQIKLSTMNNDPLSYSHYS 433
DB 339 SRLTAPLQNRMLANEREAFAVDVWGGEPLADALGALTSONLGNPFHPYEAHFYQ 398
OY 434 HPTLAER---STALD 445
DB 399 HPVPERIRVILTAD 413

RESULT 10

C69831

Conserved hypothetical protein yhnF - Bacillus subtilis

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: C69831; 140008
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehlich, S.D.; Emerson, P.T.; Ertian, K.D.; Errington, J.; Fabely, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritzt, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Manee
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli
Rieger, M.; Rivolta, C.; Kocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; M01D:98044033
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Accession: C69831
A:Molecule type: DNA
A:Residues: 1-426 <KUN>
A:Cross-references: GB:299109; GB:AL009126; NID:926333260; PIDN:CAB12869.1; PID:el183031
A:Experimental source: strain 168
R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.
Gene 70, 351-361, 1988
A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s
A:Reference number: I39994; M01D:98108019
A:Accession: 140008
A:Status: preliminary; translated from GB/EMBL/DDBT
A:Molecule type: DNA
A:Residues: 1-42; 'LESTAOA' <RES>
A:Cross-references: GB:M22914; NID:9143697; PIDN:AAA22830.1; PID:9143698
C:Genetics:
A:Gene: yhnF

Query Match 14.0%; Score 328; DB 2; Length 426;
Best Local Similarity 26.9%; Pred. No. 1.3e-17;

Matches 108; Conservative 76; Mismatches 152; Indels 66; Gaps 14;
OY 81 LAQKLVFIKIDF-----PKIMHNAVSL-----NAVLPFHHNVSTVAQSLCF 124
DB 53 VABOYNVKNFLEFVIGVPLDMFLFVLLVSGSKRIKIMIAAPFR-----LQTVGF 106
OY 125 LGLLSLSTLYDPLSYSHFVLEERKGFNKLTVQWITMIMIKSLTIAIG---PILY 181
DB 107 VEVSLTTLTVLTPLEDIGVQVSLD-YNISQTTASNAKQVISEWISFPLTCLVLFY 165
OY 182 LFLKIDKPEPTDELAMYIM-----FLFVAVQILAMTIIPEVIMPENKFTPLEDEGLKKSIE 237
DB 166 WLKIRHEKKMMLVAMLTVPFSLFLEFIQ-----VPIIDPLXNDYPLKKNLESKIL 218
OY 238 SLADRVGFPDLKIFVIDGSKRSSHNAVFTGLPFTSRIVLFDLVNSSTDEITAVLAH 297
DB 219 ELADENAPIADHYEVNMSEKTNALNAVYIGI-ANKRIYLMPTLTKLDDSELTFTMGH 277
OY 298 EIGHMOKNHIYNNMIFSQLTFLFSL-----FTSYRTSYNTSPNTEGFLKSTGSPVDP 352
DB 278 EMGHYVKKH-----VYIGLAGYLLVSLAGFVIDKIKRVIRLIRSFHLEGRDLAALP 332
OY 353 VITKEPPIIGFLFNDLITPLECAMQFVMSLISRTHEQADAYAKKL-GYKONLCRALI 411
DB 333 L-----LTLLEFVLSFAVTPFSNA-----VSRYQENKADQYGIETLENREAAVKTQ 379
OY 412 DLQIKNLSTNVPPLYSYHSHPTLAERSTALDYSEKKKN 453
DB 380 DLAVTGLSOVDPVLYKIFRGSHPISMERIQH-----AEKEN 417

RESULT 11

H75127

CAAX prenyl proteinase 1 related PAB0555 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 07-Dec-1999
C:Accession: H75127
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: H75127
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <RAW>
A:Cross-references: GB:AJ248285; GB:AL096836; NID:95458067; PIDN:CAB49737.1; PID:el151
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0555

Query Match 8.4%; Score 197; DB 2; Length 338;
Best Local Similarity 22.3%; Pred. No. 1e-07;

Matches 88; Conservative 66; Mismatches 133; Indels 108; Gaps 18;

OY 96 IMHNAVSLNAVLPVR--HMVSTVA-----QSLCFLDLSLSLTVLPLSY 142
DB 2 LMWMLI-LFGMLIPVLPVLMHRIKELRGYERKMKIKKELVLELILSLVFFSLAGILGF 60
OY 143 SHEVLEERKGFNKLTVQWITMIMIKSLTIAVIGPIL---YLFKIDKPEPTDELAMY 197
DB 61 FDFPMSTIEPLDKITVILPVLVPLVYSTLTITRELGEKIRRGDIYVAVF----- 110
OY 198 IMVFLFV---QLAMTIIPEVIMPENKFTPLEDEGLKKSIESL 239
DB 111 -IVIFLFLVLSLFTIPMPYIVYLLMFLSPALFSLIRVRFHGTGM-SGELAKAYEEL 168
OY 240 ADRVGFPLDKIFVIDGSKRSSHNAVFTGLPFTSRIVLFDLVNSSTDEITAVLAHET 299
DB 169 CRRANVAVVEEYIID---BERIGAFVTGM--KKRTIFITKGAIEKINEMELLAVIAHEL 222
OY 300 GHMOKNHIYNNMIFSQLTFLFSLFTSYRTSYNTSPNTEGFLKSTGSPVDPITKEFP 359

```

Query Match Similarity      6.5%; Score 152; DB 2; Length 291;
Best Local Similarity      24.4%; Pred. No. 0.00024;
Matches 72; Conservative   46; Mismatches 123; Indels    54; Gaps    11;

OY  179 ILYLEFLKIF--DKPEFDLWYMVFLPVVOILAMTIIPVFIMPENKFT-----PLE 228
       | : || : || : || : || : || : || : || : || : || : || : || : ||
DB   17 IASITLKILGDRFPGNVSGSLIVCAVFG--FAGSLAVFLTSKMWKMAKSTGEVISQRT 75
OY  229 DGE--LKSIIESIADRVGFPLDKIFVIDGSKRRSSISNMYFGLPFTSKRIYLFDTLVSN 286
       | : || : || : || : || : || : || : || : || : || : || : || : ||
DB   76 RHEOMLLDTVELLSSEAGIKMPEV---GIPEAYEANAFAATGMKNDAVLAVSOGILERF 131
OY  287 STDEITTAVALAEHIGHMOKNHIVNMVIFSQLHTFLIFSLEFTSYRNTSEYNFGFELEKST 346
       | || : ||||| ||||| : || : || : || : || : || : || : || : || : ||
DB   132 SPDEFKAVLAHEIGHGVANGDMVTALIQ-----GYVNFFVMEFAAIF 173
OY  347 GSPVDVPVTK--EPFIIGFM--LENDLLTPECAMQFVMSLIISTHEYQADVAKKLGY 402
       | : || : || : || : || : || : || : || : || : || : || : || : ||
DB   174 GNFDVAIKALKNEEDGGICGYFVATIIFAELV--LGILASTIVWMFSRRREFRADAGAHLAG 231
OY  403 KQNLCRAIIDIQ-----IKNLSTMNVDPLYSYHY-----SHPTLAERSTAL 444
       | : || : || : || : || : || : || : || : || : || : || : || : ||
DB   232 TGAMIAALQRLRSEQGPVOMPDTLNAPGINGLKHGLAGLLMSHPLEDRIEAL 286

RESULT  13
D71878
Stress protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Aug-1999
C:Accession: U71878
R:Aim, R.A.: Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, L.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho-
A:Reference number: A71800; MWID:99120557
A:Accession: D71878
A>Status: Preliminary

```

```

RESULT 14
C69279      heat shock protein (htpx) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Aug-1999
C:Accession: C69279
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Doi,
  : Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness,
  : Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
  : Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Attlich, P.; Kaine, B.P.; Sykes,
  : Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
  : eobacterium Pyrococcus furiosus
A:Reference number: A69250; MUID:58049343
A:Accession: C69279
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-335 <KLE>
A:Cross-references: GB:AE001090; GB:AE000782; NID:g26699413; PID:AMB90998.1; PID:g265
  : C:Superfamily: heat-shock protein htpx

Query Match      6.4%; Score 150; DB 2; Length 335;
Best Local Similarity 21.8%; Pred. No. 0.0004;
Matches 71; Conservative 51; Mismatches 113; Indels 90; Gaps 13;

OY 163 TDMIKSLTLVAIVGPIILYFLK---IFDKFPTDLMYINWFLFVQVLIANTTIIPVIMP 219
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 44 TSMVLGMLIAISAIIYLLILAVGVISF-----YGLIFLLILINLMWLLSPYIINL 98

OY 220 MKNFTPLDEDELKKSTESLADRVGFLDKLFYIDGSKRSSHNAFYTGLEPFSKRIYLF 279
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 99 SYG---AQRDERLQMVNVSARILNVKPPKAVV-----RSPPAFAFYGNFLGKFAVVS 150

OY 280 DTLVNSNTIDETIVLAIVHGMOKNHNIVNVVIFSQLTHTFISFLFSIYRNTSFYVTFG 339
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 151 ESLMRMSQDELLEAVIGHGICH--HKHRDNAV-----LLFGLLFSVI---FY--LG 195

OY 340 FLEKST-----GSFVDPVITKEPPIIGMFLFNLLTLPLECAQMEVMSIINTHEVOA 393

```

```

Db      196 YALHSSMRDRRCQAQLAIGIA--AVIVSFI-----VQILVAFSRLEXYA 241
QY      394 DAYAKKLGKYNLCRALIDL-----QI 415
Db      242 DFCGVRAITNKDAMORSIAKILHLYHRRYPDYLAPIQDSKFRLLFYAFNNAVEPTIRADI 301
QY      416 KNLSTMNVDPLYSSYHYSHTPLAER 440
Db      302 EALKMKVSPF-QEFLSTHPLPKR 325

```

RESULT 15

```

E81951
Probable membrane-bound zinc metalloproteinase (EC 3.4.24.-) NMA1031 [imported] - Neisser
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: E81951
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: E81951
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-279 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:97379424; PIDN:CAB84300.1; PID:9737973
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: htpX; NMA1031
C:Superfamily: heat-shock protein htpX
C:Keywords: hydrolase; metalloproteinase

```

```

Query Match      6.3%; Score 148.5; DB 2; Length 279;
Best Local Similarity 25.7%; Pred. No. 0.00041;
Matches 59; Conservative 42; Mismatches 88; Indels 41; Gaps 10;

QY      228 EDGELKKSIESLADRVGFPDLKIEYIDGSKRSSHNSAYFTGLPTSKRIVLEDTLVNSNS 287
Db      75 EEAWLNTVEAQAQRMNLTPEVALY---HSPEPNAPATGASRNSSLIAVSTGLDHT 130

QY      288 TDEITAVLAHIEIGHQKHHIYNN-VIFSQHTFLIESLFTSYRNTSFYNTFGEFLEKST 346
Db      131 RDEVEAVLAHEMAHANGNDMTLTLLIGVNTFVEF--LSRIANLIARNNDG--SOSQ 185

QY      347 GSFVDPVITKEFPITIGFMLFNDLTFLECAMQFVMSLSRTHFYQADAYAKKLGKYNL 406
Db      186 GTYF--LVSMVFQILFGL-----ASLIYMFPSKREYRADAGAAKLVGAPKM 231

QY      407 CRAL-----IDQIKNLSTMNV-----DPLYSYHYSHTPLAERSTAL 444
Db      232 ISALQRLKGNVDLP-EEMNAMGIAGDTRDLSLST---THPSLDNRIRAL 276

```

Search completed: February 12, 2002, 09:26:37
Job time: 63 sec

THIS PAGE BLANK (P. 3PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:37 ; Search time 26.02 Seconds
(without alignments)
922.175 Million cell updates/sec

Title: US-09-165-460A-4

Perfect score: 1651
Sequence: 1 MLOFSNPLVLYTISIVLP.....ISLKDRLQTLGVNPGYRITLV 315

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1651	100.0	315	2	SS4486
2	202.5	12.3	271	2	T37745
3	113.5	6.9	339	2	T19663
4	109	6.6	1363	2	T15653
5	104.5	6.3	255	2	B81693
6	102.5	6.2	1471	2	T40117
7	100.5	6.1	433	2	T11162
8	98	5.9	256	2	H71537
9	97.5	5.9	433	2	T17654
10	97.5	5.9	589	2	D86243
11	97	5.9	483	2	A81352
12	96.5	5.8	226	2	T09804
13	96.5	5.8	534	1	S73804
14	96	5.8	328	2	E86837
15	96	5.8	426	2	S58304
16	95	5.8	274	2	C83709
17	95	5.8	500	2	T16630
18	94.5	5.7	224	1	PMF857
19	94.5	5.7	244	2	E69787
20	94.5	5.7	269	2	T22591
21	94	5.7	226	2	T12002
22	94	5.7	346	2	T28730
23	94	5.7	347	2	D71690
24	94	5.7	400	2	T25605
25	94	5.7	769	2	S50966
26	93	5.6	645	2	T11157
27	93	5.6	2376	2	S48405
28	92.5	5.6	670	2	T43784
29	92	5.6	715	2	S10843

30	91.5	5.5	457	2	T11283	NADH dehydrogenase
31	91.5	5.5	471	2	A43956	serotonin receptor
32	90.5	5.5	492	2	E58931	NADH dehydrogenase
33	90.5	5.5	852	2	A72343	hypothetical prote
34	89.5	5.4	372	2	A70302	rod shape determin
35	89.5	5.4	390	2	T28083	hypothetical prote
36	89.5	5.4	449	2	S02011	serotonin receptor
37	89.5	5.4	471	2	A34863	serotonin receptor
38	89.5	5.4	471	2	S40689	5-hydroxytryptam
39	89.5	5.4	471	2	S11280	serotonin receptor
40	89.5	5.4	638	2	T11801	NADH dehydrogenase
41	89	5.4	347	2	S47871	NADH dehydrogenase
42	89	5.4	385	2	S43540	YSD83 protein - ye
43	89	5.4	2222	2	A37490	voltage-dependent
44	88.5	5.4	312	2	T3701	olfactory receptor
45	88.5	5.4	414	2	T19912	hypothetical prote

ALIGNMENTS

RESULT 1
S54486
probable membrane protein YMR274c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR156.1lc
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 06-Feb-1998
C:Accession: S54486
R:Lyse, G.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54014
A:Accession: S54486
A:Molecule type: DNA
A:Residues: 1-315 <LYE>
A:Cross-references: EMBL:249260; NID:g809081; PID:g809097; MIPS:YMR274C
A:Experimental source: strain AB972
A:Genetics:
A:Gene: SGD:RCCL
A:Cross-references: SGD:S0004887; MIPS:YMR274c
A:Map position: 13R
C:Keywords: transmembrane protein
F:7-23/Domain: transmembrane #status predicted <TM1>
F:45-61/Domain: transmembrane #status predicted <TM2>
F:106-122/Domain: transmembrane #status predicted <TM3>
F:210-226/Domain: transmembrane #status predicted <TM4>
F:241-257/Domain: transmembrane #status predicted <TM5>
F:280-296/Domain: transmembrane #status predicted <TM6>

Query Match	100.0%	Score 1651	DB 2	Length 315
Best Local Similarity	100.0%	Pred. No. 2e-141		
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	MLOFSNPLVLYTISIVLP	PLVATPSOPESKNDNPTIKSRMOKLTIMLSNLPVPLQ	60
DB	1	MLOFSNPLVLYTISIVLP	PLVATPSOPESKNDNPTIKSRMOKLTIMLSNLPVPLQ	60
QY	61	SQLSSTTSHISFKDAFLGII	PGYAAALPNPWOFSQFVKDKLTKCAMLLTYCGPVLD	120
DB	61	SQLSSTTSHISFKDAFLGII	PGYAAALPNPWOFSQFVKDKLTKCAMLLTYCGPVLD	120
QY	121	VLYHLNKRSSILEDPYH	HEFLNWSFRNFIAPRIETRYTSMLTYYNLTIPHSQLSQ	180
DB	121	VLYHLNKRSSILEDPYH	HEFLNWSFRNFIAPRIETRYTSMLTYYNLTIPHSQLSQ	180
QY	181	QLFMQPSLEFGLAHNAH	AYEQLOEGSMPTVSLTLTTCFOILLTYTLFGGLTKRFVPTG	240
DB	181	QLFMQPSLEFGLAHNAH	AYEQLOEGSMPTVSLTLTTCFOILLTYTLFGGLTKRFVPTG	240
QY	241	LMCCITLHALCNMGFPG	SRNLNHTVVDKRAGRISKLVSIMNKCYPALVLGLISKD	300
DB	241	LMCCITLHALCNMGFPG	SRNLNHTVVDKRAGRISKLVSIMNKCYPALVLGLISKD	300

RESULT 14
86837

RESULT 14
86837

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 : Search time 4622.84 Seconds

(Without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460a-3

Perfect score: 2948
Sequence: 1 tgaactgttgatgaacaaag.....gggagagataaagaatcaca 2948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	2948	100.0	29987	8	SC8156	249260 S.cerevisia
2	1560.4	52.9	4156	8	D50276	D50276 Saccharomyc
3	1361	46.2	4130	8	YSCNCRIG	L42821 Saccharomyc
4	1163	39.5	4225	8	SCU65682	U65682 Saccharomyc
5	1069.4	36.3	3250	8	SCU32580	U32580 Saccharomyc
6	931	31.6	3525	8	SCU63849	U63849 Saccharomyc
7	786.8	26.7	3240	8	YSCBUL1	D50083 Yeast BUL1
8	703.8	23.9	3020	8	SCRDSTGN	X88901 S.cerevisia
9	225.2	7.6	270	8	YSCGEMC8X	M87379 Yeast Eco R
10	143.8	4.9	27559	8	SC8339	249210 S.cerevisia
C 11	75	2.5	2023	8	SCYNR066C	Z71663 S.cerevisia
12	71.8	2.4	4775	8	SCU32938	U32938 Saccharomyc
C 13	68	2.3	780	11	CNS06KRB	AL402561 T7 end of
14	65.6	2.2	256172	2	AC005139	AC005139 Plasmodiu
15	65.6	2.2	310779	2	AC005140	AC005140 Plasmodiu
16	64.4	2.2	904	11	CNS06189	AL399667 T3 end of
17	64.4	2.2	86827	3	PFMAL3P5	AL034556 Plasmodiu
18	63.4	2.2	1920	3	PFGRUB	X62393 P.falciparu
19	63.4	2.2	110000	3	PFMAL13P2_0	AL049185 Plasmodiu
20	62	2.1	253305	3	PFMAL3P7	AL034559 Plasmodiu
C 21	61.6	2.1	234112	6	AX083744	AX083744 Sequence
22	60.8	2.1	318221	2	PFMAL4P2	AL035475 Plasmodiu
23	60.6	2.1	202872	2	AC016160	AC016160 Homo sapi
C 24	60	2.0	5632	3	AF310892	AF310892 Dictyoste
C 25	59.4	2.0	3214	3	AF019980	AF019980 Dictyoste
26	59.4	2.0	318221	2	PFMAL13P3	AL034576 Plasmodiu
27	58.8	2.0	321003	2	PFMAL4P3	AL034576 Plasmodiu
28	58.4	2.0	78439	17	HSUB5197	U85197 Homo sapien
C 29	58.4	2.0	169546	2	AC004157	AC004157 Plasmodiu
30	58.4	2.0	245802	2	AC006279	AC006279 Plasmodiu
31	58.4	2.0	251124	3	HUAE000660	AE000660 Homo sapi
C 32	58.2	2.0	153098	9	PFMAL3P2	AL034558 Plasmodiu
C 33	58	2.0	169794	2	AC004688	AC004688 Plasmodiu
34	58	2.0	178137	9	AC011302	AC011302 Homo sapi
C 35	57.8	2.0	13684	3	AE001403	AE001403 Plasmodiu
36	57.8	2.0	153477	2	AC006278	AC006278 Plasmodiu
37	57.6	2.0	53932	2	AC023371	AC023371 Homo sapi
C 38	57	1.9	12029	3	AE001431	AE001431 Plasmodiu
39	57	1.9	163443	2	AC006280	AC006280 Plasmodiu
C 40	57	1.9	178273	2	AC005308	AC005308 Plasmodiu
C 41	57	1.9	196149	2	AC004709	AC004709 Plasmodiu
C 42	56.8	1.9	67970	3	PFMAL1P3	AL031746 Plasmodiu
C 43	56.8	1.9	149627	2	AC087428	AC087428 Homo sapi
C 44	56.6	1.9	92633	2	PFMAL4P1_3	Continuation (4 of
C 45	56.6	1.9	153477	2	AC006278	AC006278 Plasmodiu

ALIGNMENTS

RESULT	1
SC8156/c	
LOCUS	SC8156 29987 bp DNA
DEFINITION	S.cerevisiae chromosome XIII cosmid 8156.
ACCESSION	Z49260 Z71257
VERSION	Z49260.1 GI:809081
KEYWORDS	BUL1; cytochrome b5; DAG1; initiation factor 1A; inorganic pyrophosphatase; IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF1; TFS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP; URA10.
SOURCE	baker's yeast.
ORGANISM	Saccharomyces cerevisiae
REFERENCE	1 (bases 1 to 29987)
AUTHORS	Eye, G. and Churcher, C.M.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 29987)
AUTHORS	Barrell, B. and Rajandream, M.A.
TITLE	Direct Submission

11-AUG-1997

SOANTLDGEFASNMPINNTMTWPERSLRRENTYRIRISOEKEVEQSDVEMKND
DEERKLTNTIKVEIDPHKSPROEDENMSPPSIGFODIYHROSSEEMNO
EMGIEKEAEVYKVRNDYVDEOLEREGTDMKPCATDNDKTRKRRRNGPTMKN
NMSREDNEBNOGDDENEVDSORNELDNSKHTI5LENGEKTEVSNKEENNS
TSTATSQTRKIEKTFANLFRKPHHHDSSSPSSPSSPSPINDAAVHRKSK
KLGKSGREVEPVLNRPFRHHRHSRGSOKI5VTKLDSQPOOIPLOLEBA
IEIEKRESEDESILPOLQPAVSVSTSVNSKNDREDEAKKRNKRSMTETESNOHSH
HVOKENTDEOKAOLAPAOVOYOTSVPOASAPYONAPOTSAPEVASOTOPAP
PIKHTSLIPRKILFEADVKKPDKRNSPVOTSDSFGPPLPLTYSTYIMDHRPIV
ERATIRLSHLKLSKRGLEDOVLLSNMTATYLVNHTLIMEOVADHKEQOQOQOQO
P"

BASE COUNT 1430 a 846 c 867 g 1013 t
ORIGIN

Query Match 52.9%; Score 1560.4; DB 8; Length 4156;
Best Local Similarity 99.9%; Pred. No. 4.5e-287;
Matches 1561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1387 gagcctacttgagaagatttaccatgaattcctgaataatttgaggttcaggaattt 1446
|||||
Db 1 GAGCCTTACTTGAAGATTTTACCATGAATCTCTGAATTTTGAGATTTCAGCAATTT 60
Oy 1447 tataattgcacaataacttgaggaataatttacacgtcaatgcttttgactgaactt 1506
|||||
Db 61 TATATTGCACCAATTAAGAGAAATTTTACACGTCAATGCTTTGACTGACTACTT 120
Oy 1507 aaactaataccgaacttcgcaactaagctatcaacagttatttgggaacaatcgcttt 1566
|||||
Db 121 AAACCTAATACCGCATTTGGCAACTAAGCTATCAACAGTTATTTGGCAACCATCCCTTTT 180
Oy 1567 ttgttgacttgagcagcacacacatgcttatgagaatataaggaagcttcacgaac 1626
|||||
Db 181 TTTTGACCTTGGCGACGACACCATGCTTATGACCAATTCAGAGAGCTCCATGACAC 240
Oy 1627 tgtttccattctgctgcaacaatgcttccaaatttatacaacaacacttttgagaggt 1686
|||||
Db 241 TGTTCATCTCGTGACACACATGCTTCAATTTTATACCAACACTTTTGGAGGTT 300
Oy 1687 aaccaagttgtatctgtaagaagaagcggaacactatggtgctgataatccctgcatgc 1746
|||||
Db 301 AACCAAGTTTGTATTCGTGAAGAACAGGCGGAGACCTATGCTGTCATATCTTCATATGC 360
Oy 1747 cctttgcaataatcatgaggggtttccgtgcttcaagaattgaattacattcacagta 1806
|||||
Db 361 CCTTTCGAATATCATGGGTTTCTCGTCTTCAAGATTGAATTTACATTCACAGTGT 420
Oy 1807 agacaagaagctggagcatttccaaattgctcacaatcctggaataagtgctactgc 1866
|||||
Db 421 AGACAAGAAGCTGAGCATTTCCAATTTGTCTCATCTGGAATAGTGTCTACTTGC 480
Oy 1867 actgctgctccttggaataataatccctgaagataccttacaactcgtgtagaactcc 1926
|||||
Db 481 ACTGCTGTCTTGGATTAAATATCCCTGAAGATACCTTACAACTCTGTAGGAATCTC 540
Oy 1927 tggatagaataacccttgagccttttctgtaactgtaacggttaaatctcta 1986
|||||
Db 541 TGGTATATGAATTAACCTTTTACCTTTTTCAGTACTGTATACCGGTTAAATTTCTTA 600
Oy 1987 tctactataaccttttctactactataatgaattcctacgagcagcgagctttgt 2046
|||||
Db 601 TGTACTATAACCTTTTCTACTACTATATGGAATTTATGAGAGACGGGCTTTTCTT 660
Oy 2047 acggaagagtgaaaaacgagtttggttcttggtgtaagaatttggaggaactataa 2106
|||||
Db 661 ACGGAGAGTGAAAAAAGAGTTTGTGTGTGTGTAAGAAATTTGAGAGACTATATAA 720
Oy 2107 gtacctacttggtatgaagcactaacaagtgctggtcagtgatgtaagt 2166
|||||
Db 721 GTACCTAATCTTTGTATTAACGACTCAATTAACAGTCTGTGTGCAATGATGAAGT 780
Oy 2167 tgcagactaagagtaagaagtgagcactaataagtttcagagtttcttctt 2226
|||||

Db 781 TGTCAGATCTAACAGTAGAGAGAAGTGCATCTAATAGTTTGCAGCTTTCTTTT 840
Oy 2227 taagtttttaatttggttctctagaatttaaggtcttaagtttggttgggttg 2286
|||||
Db 841 TAAAGTTTATTTTGGTCTCTAGCAATTTAAAGGCTTATAGTTTGGTTTGGTTTGG 900
Oy 2287 gttacataatttcaatccaagaaggaatttaagctgtcttcttataatgccaatagaat 2346
|||||
Db 901 GCTTACATATTTTCAATTCAAAGAGAAATTTAGCTGTCTTTTATATATCTCCAATAGAT 960
Oy 2347 aacgagacatgctgctgactacatacaagcgaataagcgatcgctgaagcgaatgaata 2406
|||||
Db 961 AACGAGACATGCTGCGTACTATACATCAAGCATTAAGCGATGCGTACGTAAGGATATA 1020
Oy 2407 cggaggttggaatttgatgctgcaagatcccttgcaatgaatccgcaaggtataa 2466
|||||
Db 1021 CGGAAGTCTGAAGTTTGTATTTGCTGACAGTCCCTTGCAATGAATCCGACGCTAAA 1080
Oy 2467 aacctaaagaattgctgattgggtlcaatlggaatttactatgtaaccagaattagata 2526
|||||
Db 1081 AACCTAAAAATTTGCTGATTTGGTCAATGATTTACTTATGATCCGAATTAATATA 1140
Oy 2527 aaattcgttggaatcctagtgaggagacatcatgctggaacacatccagttctgcg 2586
|||||
Db 1141 AAATTCGGTGGGAATTAAGTGGAGACGATCATGCTGTGACAGCATCCAGTTCTGCG 1200
Oy 2587 tcaatgccaagtgcacacacacacogttaaataacagataatagatccaactccgcta 2646
|||||
Db 1201 TCAATGCCAAGTGCACACACACCGTTAATACACACATATATGATGATCAATCCGCTA 1260
Oy 2647 gagaactgcatggaaggggtaactcaaggaatagaatccctcaataagaactaaagaat 2706
|||||
Db 1261 GAGAACTTGCAATGGAGGGGTAACCTCAGGATAGATATCTCCATTAAGACTTAACAAAGT 1320
Oy 2707 aactactaggtataaaaaaagtgctcactctccatccaggaatttaatgctaaagta 2766
|||||
Db 1321 AACTACTAGGTATATAAAAAAGTGTTCTCATCTCATCCAGGAATTAATGCTAAGCTA 1380
Oy 2767 ttaagaaaaactattataggttcccgcaataacaccccttaagcttaagcctgataat 2826
|||||
Db 1381 TTAAGAAAGAACTTATTTATTTGCTTCCCGCATATCAACCCCTTAACGTTAAACCTGATAT 1440
Oy 2827 ttccatgagcttgtaacaagatacttacaataatacaaaccaagcgaacatggtgaagat 2886
|||||
Db 1441 TTCTAGAGCTTGTAACAAGATTAATTTACAAATATATACAACTAACGCAATAGTGAAGAT 1500
Oy 2887 aatgatgggaatagcaatgaataaataacgataatgaggaataatgggaggaataa 2946
|||||
Db 1501 AATGATGGGATATGCAATGAATAATTAACGATATTGAGTATATGGGAGATTAAGAAATCA 1560
Oy 2947 ca 2948
Db 1561 CA 1562
RESULT 3
YSCNRIG LOCUS YSCNRIG 4130 bp DNA PLN 11-DEC-1996
DEFINITION Saccharomyces cerevisiae Nrc1 gene, complete cds.
ACCESSION LA2821
VERSION LA2821.1 GI:857371
KEYWORDS
SOURCE Baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1. (bases 1 to 4130)
AUTHORS Bi.E. and Pringle,J.R.
TITLES ZDS1 and ZDS2, genes whose products may regulate Cdc42p in
MOL. CELL. BIOL. 16 (10), 5264-5275 (1996)
JOURNAL 96413277
FEATURES location/Qualifiers

REFERENCE 2 (bases 1 to 4225)
 AUTHORS Jia, Y., Li, J.-P. and Butow, R.A.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUL-1996) Biochemistry, University of Texas
 Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX
 75235-9038, USA

FEATURES

source 1. 4225
 location/Qualifiers
 /organism="Saccharomyces cerevisiae"
 /strain="COP16107"
 /db_xref="taxon:4932"
 /chromosome="XIII"
 547.. 3294
 /gene="ZDS1"
 547.. 3294
 CDS
 /gene="ZDS1"
 /note="gene also called RTG2S1, CES1, NRCL; Highly charged
 protein; may be a transcription regulator; multicopy
 suppressor of mrtg2-5"
 /codon_start=1
 /product="Zds1p"
 /protein_id="AAB06617.1"
 /db_xref="GI:1508861"
 /translation="MSNRDNEMLRTTSDKALASQDRKRSEVLLIAQSLDNEIRSY
 KNLRLISGMDLIDPDLIKFGESGSRSMGTSSASMPDITLVNNTYSDP
 TPLELHGRNGSIESNKTGKQNYLGIRKGVSPSKLNNVLTMLWPAHOHN
 VKPNFLIYODITONLSDNGEDNGNSENNIDJEDNGEKSEYKKNENTIL
 NRGLSRGNASLIRPSTLRSTIEFDNEDDNKGSASSTYKVERLSKIERV
 SLRDTIELTKISAGLTNDALITLRTLSMAGSYDKDQPECHYDEGDIGEST
 SOANTLDDGFASNMPINNTWPERSLRSRNTYRINSOEKQEVDESVEMKD
 DEERLKLNTIKVEIDPHKSPFQODESNMSPGSDIFODIYNHROSSEMO
 EMGIEAEVEPVKRVNDYEDOLEREGTDMVKPSATDNETKTHRRNKGMTLN
 NKMSREDNENGDENENFVDSORMELEKHYSLINFGKKEPVSKKEENNS
 TSTSTSTQKIEKTPANLRRRKHKKHSDSSSSPSSPSPINNDAYHVRKSK
 KLGKSGREPEVPLVLRNRPFRHHHSRHSOKI SVKTLKDSQDQIPLQPLEGA
 IEIEKKESDESLSLPOLQPAVSSTSNADREBEAKKKKRSNTTETISNOHKS
 HVOKENTDEQKAOLOAPQEOVOTSPVOASAPVONASAPVOTSAVPSAQOTQAP
 PLKHTSLIPRLTFADVKRKDKNSPVOTDSAFGPPLITVSTVIMFDHRLPIV
 ERAIYRISHLKLSKRSKRLEQVLLSNFMAYILNVNHTLIMEVANDKQDQDQDQDQ
 P"

BASE COUNT 1465 a 850 c 907 g 1003 t
 ORIGIN

Query Match 39.5%; Score 1163; DB 8; Length 4225;
 Best Local Similarity 100.0%; Pred. No. 1,7e-211;
 Matches 1163; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1786 gaattacattcaagtagtagaagaagaagctgagcattcccaattgctcaat 1845
 Db 1 GAATTTACATTTCCACAGTAGTAGACACAGAACCTGGACCATTTCCACATTTGGTCTCAT 60
 QY 1846 ctggaataagtgctactcgcgactgctgctccttgatataatccctgaagatacct 1905
 Db 61 CTGGAATTAAGTGCTACTCTCGACCTGCTGCTCGATTATATCCCGAAGGACTT 120
 QY 1906 acaaaactgtaggaactcctggtatagaataaaccttaagcctttttagctactg 1965
 Db 121 ACAAACTCTGGTAGACACTCTCGTTATAGATAACCTTATAGCCTTTTATACGACTTG 180
 QY 1966 catacgtttaaaattcctctgtaactaaccctttttagctactataggaattcta 2025
 Db 181 TATACCGTTTAAATTTCTTATGTAATTAACCTTTTTCCTACTATATAGGAATTTCTA 240
 QY 2026 tcgaagcagcgagcttctgttactggaagaagtgaaataatcgaagtttgggttctgga 2085
 Db 241 TCGACGACGCGGCTTTTGTGTACGGAAGAGTGAATAAATTCGAGTTTGGTGGTGA 300
 QY 2086 aagaattggagagataaagaataacttactggtatatacggactacaatacaagtcgt 2145
 Db 301 AAGAAATTTGGAGACTATAAAGTACTCTTGTGTATATAGGACTCAATAAAGTGTGT 360
 QY 2146 tcgtctcagtgtagtgaagtgctcagatctaagaagtagagaagtgagcattcaataag 2205

Db 361 TCGGTGAGTGTAATTGAAGTTGGTGCAGATCTTAAGAGTAGAAGAGGCGCATTAATAG 420
 QY 2206 gtttcgagcttctccttctttaaaggttcttacttgctcctcctgaatttaagtcctag 2265
 Db 421 GTTTCGACGTTTCTTTTAAAGTTTAAATTTGCTCTCAATTAAGGCTTTAG 480
 QY 2266 ttgaattggttcttctgtgtggtatatacttcaattcaaaagagaatttagctgct 2325
 Db 481 TTAGTTTGGTTTGTGTTGTTGTTGTTTACATATTTTCAATTCAAGAGATTTAGCTTCT 540
 QY 2326 tttaaatgttcaatagagataacgagagcaltgctgctactacatacgaagataaagcg 2385
 Db 541 TTTATTAATGTCATATAGATTAACGAGACATGCTGCTACTACATCAAGCATTAAGCG 600
 QY 2386 atcgctagctcaagaggaataacggaagctcgaagttttagtctgctcagctccctgac 2445
 Db 601 ATCGTACTCAAAAGGATAAAGGAAAGTCAAGTTTGAATTTGCTGCACTCCCTTAC 660
 QY 2446 aatgaatcgcagcgttaaaaacttaaaagattgctgattggttcaatgattactt 2505
 Db 661 AATGAATCCGACGATTAATAAACCCTAATAAGATTGCTGATTGGTCAATGATTTACTT 720
 QY 2506 atgatacgaatagatataaaatlcggttgaggaaatcagtggaagacgatacgtgct 2565
 Db 721 ATTGATCCAGATTAAGATATAAATTCGCTGGGAACTTACTGGAGACGATCATGCTCT 780
 QY 2566 ggcagacatccagttcgtcgtcgaatgcgaatgacacacacccgttaataacacaga 2625
 Db 781 GGCAGACATCCAGTTCTGCTCAATGCGCAAGTGCACACACCCGTTAATAACACAGA 840
 QY 2626 tatagcgttccaactcgcgtgagaaacttgatagaggaggaactcaggaatagatcc 2685
 Db 841 TATAGCGATCCACACTCCCTAGAGACTTGATGGAAGGGAATCTAGGATGAATCC 900
 QY 2686 tccataaagataaacaagtgtaactacttagtataaaagaagtgctcaactccatcc 2745
 Db 901 TCCATATAGACTTAACAAGTAAGTAAGTACTAGTATTAATAAAGGTTCACTCTCCATCC 960
 QY 2746 aggaataaataagctaaactgattaaagaataactatataagtggtccgcacatcaac 2805
 Db 961 AGGAATTAATAATGCTAATGATTAAGAAAGAACTTATATGTTGCTCCGCAATCAACAC 1020
 QY 2806 cctaacgttaagcctgtaaatcttcttagagcttgtaagaacttcaaaaataataca 2865
 Db 1021 CCTAACGTTAAGCCGTGATTAATTTCTTAGAGCTTGTAAGAATCTTAAATAATACAA 1080
 QY 2866 ctaagcgaacatggtgaagaataatgataatgataatgataatgataatgataatgata 2925
 Db 1081 CTAAAGCAATGCTGAAGATTAATGATGGAATGGAATGGAATGGAATGGAATGGAAT 1140
 QY 2926 aatggagagataaagaatcaca 2948
 Db 1141 AATGGGAGGATTAAGAATCACA 1163

RESULT 5
 SCU32580
 LOCUS SCU32580 3250 bp DNA PLN 24-AUG-1995
 DEFINITION Saccharomyces cerevisiae Ces1p (CES1) gene, complete cds.
 ACCESSION U32580.1 GI:929924
 VERSION
 KEYWORDS
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 3250)
 AUTHORS Schwer, B., Liu, Y. and Shuman, S.
 TITLE Identification of CES1, a high-copy suppressor of ces1-ts mutations
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3250)
 AUTHORS Shuman, S., Liu, Y. and Schwer, B.

TITLE Direct Submission
JOURNAL Submitted (27-JUL-1995) Stewart Shuman, Sloan-Kettering Institute,
1275 York Ave., New York, NY 10021, USA

FEATURES

Location/Qualifiers

1..3250

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

466..3213

/gene="CES1"

466..3213

/gene="CES1"

/standard_name="capping enzyme suppressor"

/codon_start=1

/product="Cesip"

/protein_id="AAJ74626.1"

/db_xref="GI:929925"

CDS

translation="MSNRDNESMLRTTSSDKAIAISQDRKREVLIAQSLDNEIRSY
KRLRLSIGMDLIDPELDIKFGESGSRMSGTTSSASMPDITVNTYSDP
TPLENLHGRGNSGIESNKGNYLGKKGHSPSKNANVLLKMLVPAOHFN
VKPDLFLEVDLTQNIQSLDNGEDNGNSENNNDINDNGKESOSYENKENTINL
NRGLSRGNASLIRPSTIRSYTEPDNEDDNGKGSASTVYKVERISKIKERY
SLRDITELRTISNAGLITDNDATILKRLTSMASISDKKDPOPEGHYDEGDI
SQANTLDDEFSAMPINNTWTPERSLSRSTRNTYRIRQEQEKVEQSDVMKND
EERLKLKNTIKVEIDPKCPROQODESNMSPGIDPDIYNNHROSSEWEO
EMGIEKAEVYRVNDYEDOLEREGTDYVKPSATDNEKTRHRNRNGTWIN
NMSREDNENEGDDEENENVSORMLDMSKKHYISLENGEKTEVSNKEENNS
TSTASOTROKIEKTFANLFRKPHHKKHDSSSPSSPSPSPNDAYHVRKSK
KLGKSGREPERIVLNRPRPHHSHRHSOKISVTKLDSPOOILPOPLEGA
IEIKKESDESSELPOLAPAVSVSTKSNDRKEEKKKRRNSTETISNOQSK
HVQKENTDEQAOLOAPAEVOVTSVPVQASAPVQTSAPVBSAQTQAPAP
PLKHTSLIPRKLTFKADVKKPKDPNSPVOFTDPAFGPLPLTVSTVIMEDHRLPIV
ERAIYRSLHLKLSNKRGLREQVLISNMFAYLNLVNHITLYMEQVADHKEQQQ00000
P"

BASE COUNT 1164 a 658 c 690 g 738 t
ORIGIN

Query Match 36.3%; Score 1069.4; DB 8; Length 3250;
Best Local Similarity 99.8%; Pred. No. 1.1e-193;
Matches 1081; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1866 cactgcctgctccttgatgaataacccctgaagatacctcaaacctgtaggaactc 1925
Db 1 CACTGCTGCTCCTTGG-TTAAATATCCCTGAAGGATACCTTACAACCTCGTAGGAACATC 59
Oy 1926 ctggtatagaataaccccttagccttttaacgtactgtataccgtttaaatctc 1985
Db 60 CTGCTTATAGATAACCCCTTAGCCCTTTTACGTACTGTATACCGTTTAAATTTCT 119
Oy 1986 atgtactaaaccttttcaactactatgatgaattcatcgagcgagcggtcttctc 2045
Db 120 ATGTACTATACCTTTTTCCTACTACTATTATGAGATTCTATCGACGACGCGGCTTTTGT 179
Oy 2046 taaggagagtgaaaaaactcgagtttgcgttcttggtgtaagaatttggaagactataa 2105
Db 180 TACGGAAGAGTGAAAAATCGAGTTTGTGTTTGTGTAAGAAATTTGAGAGCATATAA 239
Oy 2106 agtacctacttgcttaacgactcaatacaagtcgcttcgtagtgtagtattgaag 2165
Db 240 AGTACTACTTGTATATAGGACTCAATACAGTGGTCTGTCGTAGTGTATGTAAG 299
Oy 2166 ttgtcaagatctaaagtagaagaagtgagcatcaataaggttcgagctttctctt 2225
Db 300 TTGTCAAGATCTAAGAGTAGAGAGAGTGCGCATCAATAGGTTTCGAGCTTTCTTTT 359
Oy 2226 ttaaggttctacttgctcctcctagaatttaagctcctagtagtatttggttcttgc 2285
Db 360 TTAAGGTTTATTTGCTCTCTAGAAATTTAAGGCTTAGTATTGTTGTTGTTTGT 419
Oy 2286 gggttaacatatttcaatccaagaagaatttagctgctcttataaagtccaatagaga 2345
Db 420 GGGTTACATATTTCATTAAGAGAGAAATTAGCTGCTTTTATTAATGCCAATAGAGA 479

Oy 2346 taacgagagcatgctgctgactacaacagcgataaggcgatcgctagccaaggatata 2405
Db 480 TTAACGAGAGCATGCTGCTGCTACTACTACATCAAGCGATTAAGCGATGCTAGTCAAGGATTA 539
Oy 2406 acggaagctcgaagtttgatctgctcgaagtcctccctbacaatgaatccgcaagataa 2465
Db 540 ACGGAGCTGATGAGTTTGTATGCTGCGACAGCTCCCTTACATGAATCCGACAGGTAA 599
Oy 2466 aaacctaaagaattgctgattgggttcaatcgatttactatlgatccagaattagatat 2525
Db 600 AAACCTAAAGAGATTGCTGATGGGTGCTCATGATTTACTTATGATCCAGATTAAGATAT 659
Oy 2526 aaattcggtgggaatcctgtagtggaagacgatacgaagctgagcagacatacagttcgc 2585
Db 660 AAATTCGTTGGGATCTTAGTGGAGACGATGCTGTGACGACGACATCCAGTTCCTGC 719
Oy 2586 gtcaatgcgaagtgcacacacacccgttaataacacagataatagcgaatccaaactcgc 2645
Db 720 GTCAATGCGCAAGTGACACACACCGTTAATATACACAGATATAGCGATCCACACTCCGT 779
Oy 2646 agagaacttcgaatggagggtaactcaaggatagaaatccctcaataagactaaagaag 2705
Db 780 AGAGAACTTGATGGAGGGGTAACTCAGGATAGAAATCCCATTAAGACTTAACAAG 839
Oy 2706 taactactgaatataaaaaaagtgctcactctccatccaggaataatgaatgctaaagt 2765
Db 840 TTACTACTTAGATATAAAAAAGCGCTTCACTCTCCATCCAGAAATTAATGCTAACT 899
Oy 2766 attaaagaaaaactatataatggtcccgcaatcaaacctcaacgttaagcctgataa 2825
Db 900 ATTAAGAAAACTATATATGAGTCCGCCCAATCAACACCTTAACGTTAAGCCTGATTA 959
Oy 2826 ttctctgaagcttgcacaaagacttcaaaaataataaactaagcgaatggtgaaga 2885
Db 960 TTCTCTAGAGCTTGACAGATACTTACAAATATACAACTAAGCGAATGTTGAAGA 1019
Oy 2886 taatgatggagataagcaatgaataacagataatgagagataatggagagataaagaatc 2945
Db 1020 TAATGATGGAGATAGCATTAAGAAATTAACGATATTGAGATATGGGAGGATTAAGATTC 1079
Oy 2946 aca 2948
Db 1080 ACA 1082

RESULT 6
SCU63849
LOCUS SCU63849 3525 bp DNA PLN 11-MAR-1997
DEFINITION Saccharomyces cerevisiae zdsip (ZDS1) gene, complete cds.
ACCESSION U63849
VERSION U63849.1 GI:1498489
KEYWORDS
SOURCE
ORGANISM baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetalia; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 3525)
AUTHORS Ma,X.J., Lu,Q. and Grunstein,M.
A search for proteins that interact genetically with histone H3 and
H4 amino terminal uncovers novel regulators of the Swel kinase in
Saccharomyces cerevisiae
Genes Dev. 10 (11), 1327-1340 (1996)
JOURNAL MEDLINE 96249381
REFERENCE 2 (bases 1 to 3525)
AUTHORS Ma,X.J. and Grunstein,M.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-1996) Biological Chemistry, Univ. of California
at Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095, USA
FEATURES
source 1..3525
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
315..3062
gene

CDS

/gene="ZDS1"
315..3062
/function="suppressor of HSL1 and HSL2; transcription
regulator"
/note="ORF YW8156.15c; CES1, see GenBank Accession Numbers
249260 and U32580; similar to Mslp encoded by GenBank
Accession Number U32938"
/codon_start=1
/product="Zds1p"
/protein_id="AAB49281.1"
/db_xref="GI:1498490"
/translation="MSNRDNEMLRTTSDKAIASQDRKRSEVLIQAQSLDNEIRSV
KRLRLSISGMDLIDELIDKEGESSGRWSGTSSASMSPTDVTNVRNPSDP
TPLENLHGNSGIESNKTGKGNVIGIKKGVASIKLWVAVNPNVHON
VKPDNLEIYODPLONTOLSDNGEDNNGNENNDIDNEKESVYENKENTITL
NRLSRHGNASLIRPSTLRSSTEFDDNDKDGSDASEYVAKVEERISIKTERV
SLRDIIEELKINSAGLTNDIAITLARLISAGSYSDKKDQDQPEGHVDEGIGEST
SOANTLDGGEFASNPINNTWPERSESLRSRPNTRYRISOEOKKEVEQSDVEMKD
DEERLKLTKNTIKVEIDPHKSPROODEENMSPSIDFODIYNHRRSGEMEO
EMGIEKAEVYPKVRNDVTEODLELBEGTDDVKKPSPATDNKEETKRHRNCGEMLN
NMSREDNEENGDDENNENVDNSRHELDNSKKHITSLNNGEKTEVSKHEMNNS
TSTATSQTRQIKETKNTANLFRKPHHKKHDSVSSPSSPSPSPINDAVHVRKSK
KLGKNSGREVEPIVLNRPRIHHRHSHRSGQKISVYTLKDSQDQDLOLEGA
IEIEKESDESSESLPOLAPAVSVSTKSNRDRDEEBAKKKRSMTTEISNOQHSK
HVOKEKNDDEOKAOLAPAOEVOYTSVONASAPVONAPVOTSAVOTAPAPAP
PLKHTSLPRLKTFADVKKPKDKPNSPVOTSDSAFGEPLITVSTVYIMEDHRLPINV
ERAIYRSLHKLNSKRGLREYVLSNFMATAYLNLVNHITLMEQVADHKDQ000000
P"

BASE COUNT 1254 a 711 c 760 g 800 t
ORIGIN

Query Match 31.6%; Score 931; DB 8; Length 3525;
Best Local Similarity 100.0%; Pred. No. 2.3e-167;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2018 gaattcctacgagcagccggtcttcttctacgagagtgagaaatcagagtttctgtgt 2077
|||||
Db 1 GAATTCATCGACGACCGGCGCTTTGTTACGGAAGATGAAANAATCGAGTTTGGGTGT 60
OY 2078 ttgttgaaagaatttggagagactataaagactatcttctgtatctacgagaccataa 2137
|||||
Db 61 TTGGTGAAGAATTGGAGAGCATATAAGTACCATCTTGTGATTCGAGACCAATATA 120
OY 2138 caagtcgtctgtcagtgatgtaagtgatcagatcgaagtagagagagtgagca 2197
|||||
Db 121 CAAGTCGTTCTGTCAGGTGATGAGTGTCAAGTCTAAGAGAGAGAGAGAGGTGCA 180
OY 2198 tctaataagttcagagtttcttcttctttaaagtttcttcttcttcttcttcttcttctt 2257
|||||
Db 181 TCTATATAGGTTTCGACGTTTCTTTTAAAGTTTATTTAGTCTCCAGAAATTTAA 240
OY 2258 ggtctcagtagtttggtttcttctgtgtgtacatattctcaattcaagaagaaatt 2317
|||||
Db 241 GGTCTAGTATGTTTGGTTTCTTTTGGGTTACATATTTTCAATTCAAGGAAATTT 300
OY 2318 agcgtctcttataatgtccaatagagataagagagcagtcgctactacatcaagcg 2377
|||||
Db 301 AGCGTCTTTTATATGTCATAGAGATTAACGAGAGCATCTCGTACTACATCAAGCG 360
OY 2378 ataaggagcagctcagtagtaaaaggagataaaggagagctcgaagtttgaattctcagagt 2437
|||||
Db 361 ATAAGGCGATCGCTAGTCAAAAGGATTAACGAGAGTCTGAGAGTTTGTATTCCTGACAGT 420
OY 2438 cccttgacaatgaaatccgcagcgtaaaaaactataaagaatgctcagatctggtcacaatg 2497
|||||
Db 421 CCCTTGACAATGAATCCGACGCTAAAAAAGATGTCGATTCGATTCGATTCGATTCGATTCG 480
OY 2498 attactatctgacccagaaatagataataaattcgtgtgggaatcagtgaggagagagat 2557
|||||
Db 481 ATTACTATTATGATCCAAATTAATATAAATTCGCGTGGGGAATCTAGTGGAGAGAGCAT 540

OY 2558 catgtctgbcagacacacccagttctgtcgtacatgccaagtgcacaacacccgttaata 2617
|||||
Db 541 CATGCTCGGACGACACATCCAGTTCTGCGTCAATGCCAAGTACCAACACCGCTTATA 600
OY 2618 aacacagatagcagatcccaactccgctagagaactctgcatggaggggttaaccagaga 2677
|||||
Db 601 ACACACGATATAGGATCCAACTCCGTAGGAACCTTGCAAGGAGGAGGTAACTCAGGGA 660
OY 2678 tagaatctcccaataagaactaaacaagaagtaactcttaggtatataaaaaagtgtaact 2737
|||||
Db 661 TAGATCTCTCAATATAGACTTAACAGAGTAACTTAAGGTATATAAANAAGGTTCACT 720
OY 2738 ctccacacaggaataataatgtaactgaactatgaagaataaacttaataatggtccgca 2797
|||||
Db 721 CTCATCCAGGAATTAATGCTTAACGATTAAGAAATAACTTTATGCGTCCGCCA 780
OY 2798 atcaaacacccttaacgcttaagccgtataatcttcctagagctgtacaagaactttaaaa 2857
|||||
Db 781 ATCAACACCCCTTAACGCTTAAGCTGATTAATTTCTAGAGCTTGTCAAGATCTTTACAAA 840
OY 2858 atatacaactaagcagacatggtgaagataatgtaggaataagcaatgaataacata 2917
|||||
Db 841 ATATCAACTTAAGCAGCAATGCTGAAGATTAATGATGGAATAGCAATAATTAACATA 900
OY 2918 ttgagataatggggagagataaagaatcaca 2948
|||||
Db 901 TTGAGGATTAATGGGAGAGATTAACATACCA 931

RESULT 7
YSCBUL1
LOCUS YSCBUL1 3240 bp DNA 10-FEB-1999
DEFINITION Yeast BUL1 gene for ORF, complete cds.
ACCESSION D50083
VERSION D50083.1 GI:773189
KEYWORDS BUL1.
SOURCE Saccharomyces cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 3240)
AUTHORS Kikuchi, Y.
TITLE Direct Submission
JOURNAL Submitted (07-APR-1995) to the DDBJ/EMBL/GenBank databases, Yoshiko
Kikuchi, Graduate School of Science, The University of Tokyo,
Department of Plant Sciences, 7-3-1 Honjo, Bunkyo-ku, Tokyo 113,
Japan (Tel:03-3812-2111(ex.4466), Fax:03-5684-9420)
REFERENCE 2 (bases 1 to 3240)
AUTHORS Yashiroda, H.
TITLE BUL1, a new component of the ubiquitination pathway in
Saccharomyces cerevisiae
JOURNAL Unpublished (1995)
REFERENCE 3 (sites)
AUTHORS Yashiroda, H., Oguchi, T., Yasuda, Y., Toh-E, A. and Kikuchi, Y.
TITLE BUL1, a new protein that binds to the Rsp5 ubiquitin ligase in
Saccharomyces cerevisiae
JOURNAL Mol. Cell. Biol. 16 (7), 3255-3263 (1996)
FEATURES
MEDLINE
FEATURES
source
1. 3240
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="13r"
181..3111
/gene="BUL1"
/codon_start=1
/protein_id="BA08787.1"
/db_xref="GI:773189"
/translation="MAKDLNDGFPKPKRNPPLRPQSDFTANSSTYMANANTRGRGR
OKOEGKGRSPSLHSPKMSIASATGILGIRPELASHASTGTCPAGGNSP
LRSTANATFVETGRSLTDDIDINNVAVLPSPFMYNTLHRHITPOGWNVDPRDHFPPSY

OEANNSATGAGSSADLSHQSISTDLAGTRSSSTNLENLIPLRTEHSHIAHOS
 AVDESLDIPILDDNDNFIDKLYTPKMTPEIITKTTHAPIPHVREES
 ILKEVTEGDLIHGFIETENKSOANLKFEWYVLESTIIDKRSRTIKRPMWD
 LSASMSYSKIALSGVDPIPADVDYDSVGLNNSRVLPGCVKTKKFFIKPLQILD
 VTCKOEHSFHCLEPPSGIDIRNNCTYSIGIKNRVIGCGHVGKSPILTNMDSDN
 LSTNYTIDARIVGDKASKLYIKEREVNYRVPFEPDANVGERMTOQLNDITKL
 VOERLDLRKIFORLEKEEPTNNDIHGADLSGTIDDSQELIQRKLDOLHIKN
 RNNYLVNNDLKGHDLDNGRSGNSGHTDTSRAMGPFVESELYKLKNSNSSSFLN
 FSHFLNSSSSSSNAGKNNDLGNKERTGLIIVKAKIPROGLPYAPSLIRKTN
 VFESKSHDQENWVRSELIPEVKKRLEKLDOLITEDNSLPHRPEIOSITTEL
 ICTTAKSDNSIPIKNSELMMNKKRLSTIALDIDRISKICEETKKNFLEINLEY
 MNNGDRPKELKFTDFTISQLENDIESICNLKVSNNLSNIRKQVSTLQKSHAL
 SEDSISHTGSSSSPSASLTPTVSSKSLFPLSGSSSTSLKFTDOIYHKWRIP
 LOYKRDIWNLERNKDIKETLIPSFESCLOCRRFYCVAMIKFENHLGVAKIDIPISVR
 OYTK"

BASE COUNT 1047 a 648 c 643 g 902 t
 ORIGIN

Query Match 26.7%; Score 786.8; DB 8; Length 3240;
 Best Local Similarity 99.7%; Pred. No. 6,1e-140;
 Matches 788; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgaactgttgatgaacaaagaagctgacaagatcaaaagcttctgacatgattcca 60
 DB 2451 TGAAGCTGTGATGAACAAAGAGAGAGCTGACAAAGCTTTGTACGATGATTTCCA 2510
 OY 61 ttcacaaatcttgtaaatctgaacaaagcttcaaaagatttcttgatataatgatc 120
 DB 2511 TTCCAAAAATTTGTGAATATGAACCAAGCTTCAACAAATTTCTTGAATTAAGAGATT 2570
 OY 121 atataataatgaatgaagagacgctgaagcacaagaagctgaattatcagattatc 180
 DB 2571 ATATAATATGAATGAGGAGACCGGACCAAGCAACTGAATTTACAGATTTTATTTAC 2630
 OY 181 ttcacagctgttcaagatcgaagaagcatttgaacttgaagtagtgcacaaact 240
 DB 2631 TTCACAGCTGTTTAAGATATGCAAGAGCATTTGCAAGATTAGTGTTCACAAACTT 2690
 OY 241 atccaaactttttaaagaaagctgacagcttaaaacaaactcaagacgcatatc 300
 DB 2691 ATCCAACTTTTAAATAAACAGGTACGATACCTTAACAAACATCAAGACGCAATTTATTC 2750
 OY 301 tgaagatcacaatcgcacacaggttaacgtagtctcgcgcgaactcagcgtcatc 360
 DB 2751 TGAGGATTCATATCGCACACAGGTAAGGTAAGTTCATCGCCCACTTCAGCGTCAT 2810
 OY 361 aacgcagtaactcctcaccagaagtagtcttctacatagcggtagctcgtctac 420
 DB 2811 AACGCAGTAACCTTCATCAAGAGAGATTATTTTACCTAGCGGTACGCTGTAC 2870
 OY 421 ttcctgaactttcagaagcagatctgtcacaagaatggtagtctgctccttaacgta 480
 DB 2871 TTCCCTGAAATTTACAGCCAGATTGTCATTAATAGGTTAGGATTCCTCTTAACAGTA 2930
 OY 481 caaacagacacataatggaacttgaacttcaataagaacataaagaacacttaattcc 540
 DB 2931 CAACAGACACATTAATGGAAGCTTGGAATTAATAGACACTTAAGAAACTTTAATTC 2990
 OY 541 aagtttgaagcctgcctatctgtaggtttctatctgcgtcgaatgaatgaattga 600
 DB 2991 AAGTTTGAAGCTGCCTATGTTAGGTTTATTTGCTTGAGTAATGATTAATTTGA 3050
 OY 601 aaacacatctggcgtagcgaagatgataccctaatctcgttggcgaagtgcacaaata 660
 DB 3051 AAACACATCTTGCGTAGGAGATTGATATCCATTTCTTTAGGCAAGTGCACAAATA 3110
 OY 661 aaaaacatgaaaaaaacctcgttacttcttatagatatagatatagtggtt 720
 DB 3111 AAAAAACATTGAAAAAATTTCTGTTACTTTTATAGATATAGATATATGATGCTC 3170
 OY 721 gctatagatgaaggtatctacgcgtcctctgtatcccatatataaatcttct 780
 DB 721 gctatagatgaaggtatctacgcgtcctctgtatcccatatataaatcttct 780

DB 3171 GCTATAGATGAGGATATTATTCGCGTCCTTGATTCCTTATTAATAAATTCCTT 3230
 OY 781 taatatgat 790
 DB 3231 TAAATGAT 3240

RESULT 8
 LOCUS SCRD1SGEN 3020 bp DNA PLN 04-JUL-1995
 DEFINITION S.cerevisiae rds1 gene.
 ACCESSION X88901
 VERSION X88901.1 GI:895906
 KEYWORDS rds1 gene; respiration deficiency suppressor.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 3020)

AUTHORS Stein,T.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3020)

AUTHORS Stein,T.
 JOURNAL Direct Submission
 TITLE Submitted (21-JUN-1995) T. Stein, Universitaet Duesseldorf,
 Botanisches Institut, Universitaetstr.1, D-40225 Duesseldorf, FRG

COMMENT Related sequence D50083.
 FEATURES
 source
 location/Ouallifers
 1..3020
 /organism="Saccharomyces cerevisiae"
 /strain="GRF8"
 /db_xref="taxon:4932"
 /clone_lib="YCp50 genomic genebank"
 /clone="pUC19/Yep352"
 32..2962
 /gene="rds1"
 32..2962
 /gene="rds1"
 /codon_start=1
 /product="respiration deficiency suppressor"
 /protein_id="CA61363.1"
 /db_xref="GI:895907"
 /translation="MAKDLNDGFPPEPRKRLRPDRSPFANSSTMMNVNANTRGGR
 OKOEGKSSRSPLSIPKSPKSWISASATGILGRLPILASHABSTGPPACNSP
 LRSTANATPEVGRSLTDDIDNNVDVLPSEFMYNTLHHITGCVADPDRHPPSY
 OEANNSATGAGSSADLSHQSISTDLAGTRSSSTNLENLIPLRTEHSHIAHOS
 AVDESLDIPILDDNDNFIDKLYTPKMTPEIITKTTHAPIPHVREES
 ILKEVTEGDLIHGFIETENKSOANLKFEWYVLESTIIDKRSRTIKRPMWD
 LSASMSYSKIALSGVDPIPADVDYDSVGLNNSRVLPGCVKTKKFFIKPLQILD
 VTCKOEHSFHCLEPPSGIDIRNNCTYSIGIKNRVIGCGHVGKSPILTNMDSDN
 LSTNYTIDARIVGDKASKLYIKEREVNYRVPFEPDANVGERMTOQLNDITKL
 VOERLDLRKIFORLEKEEPTNNDIHGADLSGTIDDSQELIQRKLDOLHIKN
 RNNYLVNNDLKGHDLDNGRSGNSGHTDTSRAMGPFVESELYKLKNSNSSSFLN
 FSHFLNSSSSSSNAGKNNDLGNKERTGLIIVKAKIPROGLPYAPSLIRKTN
 VFESKSHDQENWVRSELIPEVKKRLEKLDOLITEDNSLPHRPEIOSITTEL
 ICTTAKSDNSIPIKNSELMMNKKRLSTIALDIDRISKICEETKKNFLEINLEY
 MNNGDRPKELKFTDFTISQLENDIESICNLKVSNNLSNIRKQVSTLQKSHAL
 SEDSISHTGSSSSPSASLTPTVSSKSLFPLSGSSSTSLKFTDOIYHKWRIP
 LOYKRDIWNLERNKDIKETLIPSFESCLOCRRFYCVAMIKFENHLGVAKIDIPISVR
 OYTK"

BASE COUNT 982 a 611 c 598 g 829 t
 ORIGIN

Query Match 23.9%; Score 703.8; DB 8; Length 3020;
 Best Local Similarity 99.6%; Pred. No. 3,7e-124;
 Matches 716; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 1 tgaactgttgatgaacaaagaagctgacaagatcaaaagcttctgacatgattcca 60
 DB 2302 TGAAGCTGTGATGAACAAAGAGAGAGCTGACAAAGCTTTGTACGATGATTTCCA 2361


```

Oy 61 ttcaaaatttgtagaata tgaaccaagttcaacaagaatttcttgatataatgaagt 120
Db 2362 TTCAAAATTTGTGAATATGAACAAGTTCACAGAAATTTCTTGATTAATAGAGTT 2421
Oy 121 atataatgaatgaaggagacgtaggcacaaaggaactgaattacagatttactaac 180
Db 2422 ATATATATATGAATAGGAGAGACCGTAGGSCAAAGGAACTGAATTTACAGATTTATTTAC 2481
Oy 181 ttcaagctgttttaacgattatcgaaagcatttgcaattgaagttaagtgttcaaacatt 240
Db 2482 TTCACAGCTGTGTTACGATATCGAATCGAATTTGCAACTTGAAGTTAGTGTTCACAACTT 2541
Oy 241 atccacaatttttaaaaaaacagtgtagtaccctaaacaacactaaagcaacgcatatc 300
Db 2542 ATCCAAACATTTTAAAAAACAGTCAAGTACCTTAACACACACTCAAAAGCAGCATTTATC 2601
Oy 301 tgaagattcaatatacgcacacaggaaggaagtagtcatcgtgcgccagttcaagcgcatc 360
Db 2602 TGAGGATTCATATATCGCACACAGGTAACGTAAGTTCATCGCCAGTTCAGCGTCAAT 2661
Oy 361 aagccagtagtacttctcatcacaagtagttatttttactaagcgttagctgtctac 420
Db 2662 AACGCCAGTAACTTCTTATCCAAAGATGATTTATTTTACCTACCGGTAGCTCGTCTAC 2721
Oy 421 ttccctgaatttacaagaccagattgttcataaatggttagagattgtctccttacaagta 480
Db 2722 TTCCCTGAATTTTACAGACAGATGTTCATTAATAGGTTAGAGATTGCTCCTTTACAGTA 2781
Oy 481 caaacagagacattaatgtgaacttgaatttaataagacattaaagaaactttaattcc 540
Db 2782 CAACAGACATATTAATGATGAACTTGGAATTTAATAGACATTAAGCAAACTTTAATTC 2841
Oy 541 aagtttaaaagtgccctgattgttaggttttatgtcgttcgtagaagatgaatttga 600
Db 2842 AAGTTTAAAGCTGCTATGTTGATGTTTATGCTTGAGTGAATGAATTAATTTGA 2901
Oy 601 aaaccatttgtaggaagattgataccctattctgttagcaagtagacaat- 659
Db 2902 AAACCATCTTGCGGTAGCAGGATGATATCCCTATTTCTGTAGCAAGTACAAATA 2961
Oy 660 aaaaaaacattagaataaattctcgttacttcttcttataagatatagatatgtatg 718
Db 2962 AAAAAAACATTAAGAAAAATTTCTGTTACTTTTATATATATATATATATATATGCT 3020

RESULT 9
YSCGENC58X 270 bp DNA PLN 11-MAR-1996
LOCUS Yeast Eco RI fragment.
DEFINITION M87379
ACCESSION M87379.1 GI:1220615
VERSION
KEYWORDS Saccharomyces cerevisiae DNA.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 270)
AUTHORS Davies,C.J., Trgovcich,J. and Hutchison,C.A. III.
TITLE Limited application of a random DNA sequencing strategy to the
JOURNAL Saccharomyces cerevisiae genome identifies 3 new genes and
FEATURES indicates that approximately 90% of the genome may be transcribed
Unpublished (1994)
location/Qualifiers
source 1..270
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
BASE COUNT 67 a 30 c 71 g 102 t
ORIGIN
Query Match 7.6%; Score 225.2; DB 8; Length 270;
Best Local Similarity 94.1%; Pred. No. 4e-33;
Matches 256; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

```

```

Oy 2018 gaattctacgagcg-accggagctttgttaccgaaagatgaaaaatcagatttggtg 2076
Db 1 GAATTCATTCAGAGACACCGGATTTTGTATCGAAGAGTCAAAA--TCGCGTTTGGTG 58
Oy 2077 ttgtgtgaagaatttgaagactataaagtaacctacttctgtattacgactacaata 2136
Db 59 TTTGTGTAAAGAAATTTGAGAGACTATAAAGTACTATACCTTGTATTAACGGGGTCAATA 118
Oy 2137 acaagctgtcgtgtcagatgtgatataagattgtgcagatcctaaagtagaagaagtgcc 2196
Db 119 ACAAGTCTGTGTCAGTGTGATTAAGATTGTCTAGATCTAAGAGTAGAGAAAGCTGCG 178
Oy 2197 atctaatagttcagagctttcttctttaaagtttttatttgctccatagaatta 2256
Db 179 ATCTAATAGTTTTCACAGCTTTTCTTTTATTTTAAAGCTTTTATTTGCTCTCTAGAAATTA 238
Oy 2257 aggtcttagttagtttgttctgttctgtg 2288
Db 239 AGGCTTAGTTAGTTTGTGTTTGTGTTGTG 270

RESULT 10
SC8339 27559 bp DNA PLN 11-AUG-1997
LOCUS S. cerevisiae chromosome XIII cosmid 8339.
DEFINITION Z49210 Z71257
ACCESSION Z49210.1 GI:798881
VERSION
KEYWORDS aminoglycoside resistance; ATR1; DAT1; MDH1;
oligo(dA)/oligo(dT)-binding; orotate phosphoribosyltransferase;
SEC65; signal recognition particle; SNO1; spore germination; URA5;
VAN1; vanadate resistance.
baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 27559)
AUTHORS Skelton,J. and Churcher,C.M.
TITLE Unpublished
JOURNAL 2 (bases 1 to 27559)
AUTHORS Barrell,B., Rajandream,M.A. and Walsh,S.V.
TITLE Direct Submission
JOURNAL submitted (02-MAY-1995) Saccharomyces cerevisiae chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT
Notes:
All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis.
Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)
is given for each CDS.
Cosmid 8339 is overlapped at the start by lambda 7056 and at the
end by cosmid 9514, EMBL X80835.
location/Qualifiers
source 1..27559
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8339"
/map="13L"
<1..1103
/note="YM8339.01, unknown, len: 366, CAI: 0.11, overlaps
and extends YM7056.08"
/codon_start=3
/product="unknown"
/protein_id="CA89100.1"
/db_xref="GI:798882"
/db_xref="SWISS-PROT:Q03210"
CDS

```

[illegible]

ACCESSION	U32938
VERSION	U32938.1
KEYWORDS	GI:1236326
SOURCE	
ORGANISM	baker's yeast. Saccharomyces cerevisiae
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
AUTHORS	Yu,Y., Jiang,Y.W., Wellinger,R.J., Carlson,K., Roberts,J.M. and Stillman,D.J.
TITLE	Mutations in the homologous ZDS1 and ZDS2 genes affect cell cycle progression
JOURNAL	Mol. Cell. Biol. 16 (10), 5254-5263 (1996)
MEDLINE	9641376
REFERENCE	2 (bases 1 to 4775)
AUTHORS	Stillman,D.J.
TITLE	Direct Submission
JOURNAL	Submitted (01-AUG-1995) David J. Stillman, Division of Molecular

COMMENT	On Mar 27, 1996 this sequence version replaced gi:987502
FEATURES	Location/Qualifiers
SOURCE	1. .4775

```

gene
CDS
    /organism="Saccharomyces cerevisiae"
    /db_xref="taxon:4932"
    /chromosome="XIII"
    /map="13L"
    669..3437
    /gene="DJS2"
    669..3437

```

`/gene="ZDS2"`
`/note="multicopy suppressor of a sin4 defect; similar to`
`yeast genes NRC1 encoded by GenBank accession number`
`L4281 and CES1 encoded by GenBank accession number`
`U32580. Method: conceptual translation supplied by author`
`/date="2001-01-01"`

```

/codon_start=1
/product="Zds2p"
/protein_id="AAB37541.1"
/db_xref="GI:1236327"
/translation="MVLMEDEVONKDHNTVESSGSDNNNNIOARRRKRQLSKKEI
FEKRSVDVLTAKSLDFTETONVKKIKRSLSTSMGVITDPLFEKNSNGSSNSQSK

```

SLOEHLHENITRFOKEEGSENDAYEEGDATVNDSDITOTYEYLAHDETELEKEE
 ITIRNASSTESARATYRNRRLSGVITLAHVLEVLENDHNSKAMVDLÖNLWLPADDE
 HPNKPENYVELIÖDTLONIQISTMODIDENKLEGVNHHVSNRKRSTVVRSPRPADE
 TSYTFDEPPLADKPÖCEQIÖVDRIRISSDITKITRVSLEKIEITBELKIRINAGATIT
 SDAYIVLASISMSGFTNESHILNÖHNTHEENEFNEENGFLIPRPSIRSRKFE

TYKILLESSILPQAVKLANSLMNIOQINDKRRASPAASTIYQYPOEASLNDTHEITFDTH
RTSDTQWSTENKGYDSPNYISDEEDLTHASISQSSSLTSDSSNNVYLIPHTGSA
ISEKIDQWSEKSNKNTNRRSEINHGWSMLANSNGSLNMGPTQOULTDDEDECVDG
EKADENVLSVSRRAKSTKASERINHKSNNRSPFIQISEKASVJLTPDVSSESQ
PSKPTPAAYEKKVELPTDITQASTHKKSLLEKRLAKLEKRRQHSQSTCSVDVYIKKST

1616 017 1000

Count	1013 d	688 c	94 / g	1325 f
Deery Match				
Local Similarity	2.48;	Score 71.8;	DB 8;	Length 4775;
Local Similarity	71.88;	pred NC 0.00050;		

	94:	Conservative	0:	Mismatches	37:	Indels	0:	Gaps	0:
2401	gataaacggaagctcgtgaagtattgatattggctgcacgcccttcgacaatgaatccgcagc	2460							
804	gagaaggagaaattcgacctgttaatttgcgcacactcttttgatnctctaatttcagat	863							

2461 gtaaaaccctaaaagatgtcgaattggtgcacfgatctactatgatccagaatta 2520
|| ||||| | | ||||||||| ||||||||| | ||||| || ||
864 gtcaaaatttgtagcgattgtcattggttcattgcatattgccctgacctga 923

QY	2521	gataataaatt	2531
Db	924	GAGTCAAGT	934

RESULT	13
CNS06K8B/c	
LOCUS	CNS06K8B
DEFINITION	780 bp DNA SRS 10-JAN-2001
ACCESSION	U7 end of clone AT0AA002H08 of library AT0AA from strain CBS 43111
	of <i>Saccharomyces servazzii</i> , sequence tagged site.
	AT000651

ACCESSION	AL402561	GI:12161691
VERSION	AL402561.1	
KEYWORDS	STS.	
SOURCE	<i>Saccharomyces servazii</i> .	
ORGANISM	<i>Saccharomyces servazii</i> .	
	Pharvotri. Fungi. Ascomycota. Saccharomycotina. Saccharomycetes.	

REFERENCE

1 (bases 1 to 780)

AUTHORS

Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H., Artiguenave,F., Mincker,P., and Galliard,J.C.

TITLE

Genomic Exploration of the Hemiascomycetous Yeasts: 7.

JOURNAL
PUBMED
REFERENCE
AUTHORS

Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)
1152882
2 (bases 1 to 780)
Souclet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bojotin-Fukuhara, M., Bon, E., Brothier, P., Casargola, S., de-Montigny, D., Dujon, B., Durieux, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogiropoulos, O., Potter, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Weslowski-Louvel, M., Wincker, P. and Weissenbach, J.

TITLE	Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)
PUBMED	11152876
REFERENCE	3 (bases 1 to 780)
AUTHORS	Genoscope

TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage
2 rue Gaston Creteilux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqlife@genoscope.cns.fr Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen

yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*

the other extremity of this insect.

Location/Qualifiers

```
source
I..70
/organism="Saccharomyces servazzii"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0AA002H08"
/clone_1ib="AT0AA"
```

```
misc_feature
/feature_id="17"
complement(<256..>780)
/note="similar to Saccharomyces cerevisiae ORF YMR275c [
BUI1; ubiquitination pathway protein ]"
/evidence="not experimental"
```

BASE COUNT	251	110 c	138 g	255 t	26 others
misc_feature	complement(<65, >771) /note="similar to Saccharomyces cerevisiae ORF YLL11w [strong similarity to ubiquitinacation protein Bulp]" /evidence=not_experimental				

Query Match	2.38;	Score 68;	DB 11;	Length 780;
Best Local Similarity	56.2%;	Pred. No. 0.0032;		

	Query March	2.2%	Score 65.6	DB 2,	Length 256172;
	Best Local Similarity	47.2%	Pred. No. 0.0085;		
	Matches 232; Conservative	0;	Mismatches 259;	Indels 1;	Gaps 1;
Oy	466	tgcctcctaagtaacagaacatattgtagtacttggaattataaaggacattaa	525		

RESULT	15
LOCUS	AC005140
DEFINITION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.
ACCESSION	AC005140
VERSION	AC005140.8 GI:9797735
KEYWORDS	HTG: HTGS_PHASE1.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Bekaryoti, Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 310779)
TITLE	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T., Kirdi,O.B., Conway,A.B. and Davis,R.W.
JOURNAL	Plasmodium falciparum 3D7 chromosome 12 Unpublished
REFERENCE	2 (bases 1 to 310779)
AUTHORS	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE	Direct Submission
JOURNAL	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT	On Aug 12, 2000 this sequence version replaced gi:8810455.
NOTE:	This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
	This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
	1 187308: contig of 187308 bp in length
	* 187309 187508: gap of unknown length
	* 187509 257820: contig of 70312 bp in length
	* 257821 258020: gap of unknown length
	* 258021 307255: contig of 49235 bp in length
	* 307256 307455: gap of unknown length
	* 310775 310779: contig of 3324 bp in length.

FEATURES
source

Location/Qualifiers
1. .310779
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PFYACB8-420"
/clone="3D7"

BASE COUNT 126134 a 29409 c 30056 g 124579 t 601 others
ORIGIN

Query Match

2.28; Score 65.6; DB 2; Length 310779;
Best Local Similarity .47.2%; Pred. No. 0.0084;

Matches 233; Conservative 0; Mismatches 259; Indels 1; Gaps 1;

```
QY 466 tgcctcctttagacgaacgaacgacatlaagtgaaacttggaatttaagaacatlaa 525
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87264 TTCTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 87323
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 526 ggaaccttaattccaaagtttgaagctgcctagtgtaggttatttcgttcgagt 585
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87324 ATATTTAAATTTATTAATGCTTTATTAATTAATTTATTAATTAATTAATTAAT 87383
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 586 aatgattcaaatctgaaacacatctggcgaagatgatacccatcttcgttag 645
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87384 AATTATATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 87443
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 646 gcaagtgacaaataaaacacatlagaaaaatccctgtaacttctctatagatag 705
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87444 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 87503
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 706 atatatgtagtcttctatagatgaagatattatcgcgctccttgatcccatla 765
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87504 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 87563
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 766 ttaataaaatcttctaaatgcaatctcggcgctcctcttgctcgtcgtatctt 825
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87564 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 87623
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 826 ttcttgaccacatcgatlgaaacacatcttgaatcttattacatttlaagttact 885
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87624 TTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 87683
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 886 aaaaatcgaagat-ttcaaggaacaaacatagaaatcttcttgcaagaataaaacg 944
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87684 TTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 87743
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 945 aaataaattgat 956
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87744 AAATTAATTAATTAAT 87755
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: February 12, 2002, 13:28:12
Job time: 10200 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460a-1
Perfect score: 1825
Sequence: 1 accacaccttttttctatct.....aaaataaacgattaacatt 1825

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pl:*
10: gb_ro:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_om:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_inv:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_inv:*
36: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1823.4	99.9	1825	8	SCYR117W	249617 S. cerevisia
2	1553.4	85.1	1706	8	SCYR117W	077137 Saccharomyc
3	342	18.7	1555	8	SCYR116W	249616 S. cerevisia
4	195	10.7	99360	8	NCB11N2	AL513444 Neurospor
5	187	10.2	36920	8	SPAC3H1	268144 S. pombe ch
6	137.2	7.5	1275	8	AF353722	AF353722 Arabidops
7	136	7.5	686	6	AB6270	AB6270 Sequence 92
8	136	7.5	686	6	ARI55763	ARI55763 Sequence
9	136	7.5	686	6	E66288	E66288 Genome DNA
10	121	6.6	935	8	SCYR118C	249618 S. cerevisia
11	117.4	6.4	1859	9	AF064867	AF064867 Homo sapi
12	117.4	6.4	2963	9	AB016068	AB016068 Homo sapi
13	117.4	6.4	2963	9	HSY13834	Y13834 Homo sapien
14	117.4	6.4	2968	6	E32056	E32056 Human AF01
15	114.2	6.3	1340	3	AF358443	AF358443 Physarum
16	98.4	5.4	78592	3	AC004335	AC004335 Drosophila
17	98.4	5.4	102413	2	AC020024	AC020024 Drosophila
18	92.8	5.1	41327	3	CBR647J19	CBR647J19
19	91.6	4.8	40767	3	CEC04F12	CEC04F12
20	88	4.8	317511	1	CJ11168X3	AL139076 Campyloba
21	86.6	4.7	582	8	CNS018P	AL10644 Botrytis
22	81.6	4.5	660	9	CNS01AH3	AL112927 Botrytis
23	72.2	4.0	2161	9	AK027874	AK027874 Homo sapi
24	66.8	3.7	1072	6	AX053334	AX053334 Sequence
25	66.6	3.6	124808	2	AP000652	AP000652 Homo sapi
26	65.6	3.6	16488	1	AE004992	AE004992 Halobacte
27	65	3.6	256172	2	AC005139	AC005139 Plasmodiu
28	64.4	3.5	169423	2	AC069309	AC069309 Mus muscu
29	64.4	3.5	186306	2	AC083946	AC083946 Mus muscu
30	63.6	3.5	1245	6	AX055202	AX055202 Sequence
31	63.6	3.5	19715	1	AE000555	AE000555 Helicobac
32	63.6	3.5	234112	3	PFMAL4P2	AL035475 Plasmodiu
33	62.4	3.4	13363	3	AE001374	AE001374 Plasmodiu
34	61.8	3.4	98734	2	PFMAL1P2	AL031745 Plasmodiu
35	61.6	3.4	14943	1	AE001528	AE001528 Helicobac
36	59.8	3.3	12029	3	AE001392	AE001392 Plasmodiu
37	59.8	3.3	110000	2	AC087899	AC087899 Mus muscu
38	59.2	3.2	114736	3	PFMAL3P3	298547 Plasmodium
39	59.2	3.2	253305	3	PFMAL3P7	AL034559 Plasmodiu
40	58.8	3.2	98899	2	AL603743	AL603743 Danio rer
41	58.8	3.2	318221	2	PFMAL13P3	AL049184 Plasmodiu
42	58.8	3.2	318503	2	AC074166	AC074166 Mus muscu
43	58.4	3.2	172239	2	AC069587	AC069587 Homo sapi
44	58.2	3.2	12029	3	AE001412	AE001412 Plasmodiu
45	58.2	3.2	187469	9	AL357312	AL357312 Human DNA

ALIGNMENTS

RESULT 1
SCYR117W
LOCUS SCYR117W 1825 bp DNA 11-AUG-1997
DEFINITION S. cerevisiae chromosome X reading frame ORF YCR117W.
ACCESSION Z49617 Y13136
VERSION Z49617.1 GI:1015836
KEYWORDS
SOURCE
ORGANISM
baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
REFERENCE
1 (bases 1 to 1825)
AUTHORS Rose, M., Koetter, P. and Entian, K.D.
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1825)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

Martinsried, FRG; E-mail: Mewes@mpi.embnet.org

FEATURES
Location/Qualifiers

source

1. 1825

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="X"

gene

343..1704

/gene="STE24"

343..1704

/note="ORF YJR117w"

CDS

/protein_id="CA89647.1"

/db_xref="GI:1015837"

/db_xref="SWISS-PROT:P47154"

translation="MPDLKTLIDHPNIPMKILISGFSAQSFSESYLTROYKLSER
KLPLVLEIDEDIFHRSKNTSRKAKSISIGDYNNLQKLVFKYIDLPKIMHAYS
LLNVLVPRFMVSTVAOSLCFLGLSSLSLVLPVLSYHFLVEEFGFKLTVOL
WITDMIKSLTALAIVGIPILYFLPKIDEPDPLWIMVLFVVOILAMTIIFFVIM
PMFNKFTPLEDEGLKSIKSLADRVGFLDKIPIYIDGSRSSHSNAYGTIPFTSKRI
VLFDPLVNSSTDEITAVLAIEIGHMOKNIIVNNVIFISQLHTFLIFSLFTSNTSR
VNTGFELEKSTGSEFVDPIKEPPIITGEMLDITPLECAMQFVMSLSRTHENO
ADAYAKKIKYKONICRALIDQIKNLSTMNDPLYSSITHSHPTLAEHLTADLVSEK
KKN"

BASE COUNT 570 a 397 c 280 g 578 t

ORIGIN

Query Match 99.9% Score 1823.4; DB 8; Length 1825;

Best Local Similarity 99.9%; Freq. No. 0;

Matches 1824; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 accacattttttctatcttcaacaagaagccttacacacacacacacatc 60
|
|
|
Db 1 ACCTACCTTTTCTATCTTCAACAGAAAGCCTTACACACACACATCATCT 60
|
Oy 61 acatacatatatacatatataatgtaaaactgtatatacttctataacca 120
|
|
|
Db 61 ACATACATACATACATAATATACATATATGTAACCTGTATATTCCTATTACCAAA 120
|
Oy 121 aagaggaatttaacttttccctcttttctacgtcatttactcaaaaactta 180
|
|
|
Db 121 AAGAGGCAATTAACCTTTTCCCTCTTTTCTACGTCATTTACTCAAAAACCTTAATTCCT 180
|
Oy 181 tgcgtctgcttcgcatcttctccagaaaaaatcgacggaataaaaaaaag 240
|
|
|
Db 181 TCGTCTGCTTCTGCGCATTTTCTCCAGAAAAAATCGACGGCAATTAATAAAGAC 240
|
Oy 241 aacgaacaagaagaaatltcggaatatataaacactctataaataacagga 300
|
|
|
Db 241 AACGAACAAGAAAGAAAGTCCGGAATTAATAACCACTTCTATTAATTAACGAAGAAGA 300
|
Oy 301 aggaataaagaagaaatgaaatcgagcgttctatctatctgattgattta 360
|
|
|
Db 301 AGGAATAAAGAAAGAAATAGAAAACTGCAGCCTTATTCATCTTTGATCTTAAGACG 360
|
Oy 361 attctcgacatcccaatataccgctggaatatcaatcatcttctgggttc 420
|
|
|
Db 361 ATTCTCGACATCTTAATATCCCGTGAATTAATCATCTTCTGGGTTCTCGATTCGCCAA 420
|
Oy 421 ttctcttcgaattctacttgacgtacagacagtlaccagaagctctctga 480
|
|
|
Db 421 TTTCTTTTCGAATCTTACTTGACGTACAGACGTACAGAAAGCTTCTGAAACAAGTTG 480
|
Oy 481 ccaactgtcgtggaagaagaaatgaaatgaaatcttcatataatcaagaact 540
|
|
|
Db 481 CCACCTGTGCTGGAAGAGAAATGATGATGAATCTTTCAATAAATCAAGGAACATCTCC 540
|
Oy 541 cgggccaaggcgaagttctcaatttgggtgagtcataaccgtgccaagaag 600
|
|
|
Db 541 CGGGCCAAGGCGCAATCTTCCATTTTGGTGAGCTCTATTAACCTAGCCCAAAAGCTAGT 600
|
Oy 601 ttcatcaatacgaacctctccctaataatctgacacatgagccgttctct 660
|
|
|

|||||
Db 601 TTCACTAATACGACCTCTCCCTTAATATGCGCAATGCGCGCTTCTTATGATGCA 660
|
|
|
Oy 661 gtcctcgacatgacatltcatatggtctccactgtcgcagagttatgctcttgg 720
|
|
|
Db 661 GTCCCTGCGCAGTCAATATGATGCTCCACTGCGCACAGAGTTATGCTTGGGT 720
|
Oy 721 ccttaccagttgttaccctgtgtattggtgcaaccttctactaagcaatttgc 780
|
|
|
Db 721 CTCTTATCCAGTTTGTCTACCTTGCTGATTTGACCTCTCTACTATAGCCATTTGTC 780
|
Oy 781 ctggaagaanaatltglttcaataaattgacccgtccaaactatgataccagat 840
|
|
|
Db 781 CTGGAAGAAATTTGTTCAATTAATTTGACCGTCCAACTATGATACCGATTTGATC 840
|
Oy 841 aagagctgaacttggggtatgataatgttggcccaatcccttaccgttcccta 900
|
|
|
Db 841 AAGACTCTGACTTGGGCTATGCTATGCTGCGCCAAATCCTTACCTTCTTAAAGATC 900
|
Oy 901 ttgataaattccctactgatttcccttggtaacatatagtctcttctgttgc 960
|
|
|
Db 901 TTTGATTAATTCCTACTGATTTCTCTTGTATCATTTATGCTCTTCTTGTGTTGCCAA 960
|
Oy 961 attctagccaatgacaatcattccagltcttcaatcattgcccatttataa 1020
|
|
|
Db 961 ATCTTAGCCATGACAAATCATTCACAGTCTTCATCATGCCCATGTTAATAAGTTCAC 1020
|
Oy 1021 ttgagagacgttgaactgaaataaataatctatgaagttggccgataagttgg 1080
|
|
|
Db 1021 TTGAGGAGCGGTGAACCTAATAAATCTATTTGAATTTGGCGAGATGAGTTGGTCCCT 1080
|
Oy 1081 ctagaataagatttgttcatcttgaagcgtccaaagaatcttctcatcaaga 1140
|
|
|
Db 1081 CTAGTAAAGATTTTGTCTATTTGACGCGCTCAAAAGATTTCTCATTTCAACGATATTTC 1140
|
Oy 1141 acaaglttgcattccactccaagaagaatltgttlttgcgaactttagtga 1200
|
|
|
Db 1141 ACAGSTTTGCCATTCACCTCCAAAGAAATGTTTGTGTCGACACTTATAGTGAACGTAT 1200
|
Oy 1201 tctactgtatgaataatgaagcgttlttggcccaatgaatcggtcaactgtg 1260
|
|
|
Db 1201 TCTACTGATGAATAATTAACGGCTGTTTGGCCCATGAATCGGTCACCTGGCAAAAACAC 1260
|
Oy 1261 atcgttaataatgcatcttctgaatcattgcaacactcccaatttccctt 1320
|
|
|
Db 1261 ATCGTTAATATGTCATCTTTATGTAATTTGCAACACTTCCATTTTCTCCCTTTTCCAC 1320
|
Oy 1321 agcaatcacaagaatatacatlttcaacaacacttgcgcttlttcttaga 1380
|
|
|
Db 1321 AGCATCTACAGAAATATCATATTTTAACAACACTTGGCTTTTCTTGTAGAGATCCACT 1380
|
Oy 1381 ggcagtttgttgaatcccgatataatcgaagaatcccaattatcatatg 1440
|
|
|
Db 1381 GGCAGTTTGTGATCCCGTATATCATTAAGGAATTCCTCATTAATGATTAATGTTA 1440
|
Oy 1441 tttaacgactatataacccaactcgaaatgtccatgcaatcgtaagactt 1500
|
|
|
Db 1441 TTTAACGACTTATTAATCTCCACTCGAATGTGCCATGCAATTCGTATGATTAATTTCC 1500
|
Oy 1501 agaaatcagaatatacaagctgtatgtcttaaaaaaatttgggttacaaga 1560
|
|
|
Db 1501 AGAATCTATGAATATCAAGCTGATGCTTATGCTTAATAAATTTGGGCTTACAGCAAAATCTA 1560
|
Oy 1561 tgaaggtctcaatgtatctacaatcaaaaacttccaccaatgaatgtatg 1620
|
|
|
Db 1561 TGTAGGCGCTCTTAATTTGATCTCAATCAAAATCAAAATCTTCCACCATGAATGTGATCTCTG 1620
|
Oy 1621 tattctagctatcatatttccatcccaacttaactgtgaatatgacgcgtt 1680
|
|
|
Db 1621 TATTCTACTATCATATTATTCATTCATCAACTTACTGTAAGAAATTTGACCGCTCTGACATAT 1680
|
Oy 1681 gttatgtaagaagaagaaataatctataagatcacatattagatgacctt 1740
|
|
|

Accession	Sequence	Position
Db 1681	GTTCGTGAAAAGAAAGAAAACCTAATCTATAGAGTCACATATTACCATGTCCGCTTAAT	1740
Oy 1741	tcgaattccgttatgtcctatatcacacaaacaaagatcctactataagaataaagga	1800
Db 1741	TCGACTTCGTTATGTCTATATCTTACCTACATACATACACAGTATCTACTATAAGATTAAGga	1800
Oy 1801	aagaaaaataaacgattaaacatt	1825
Db 1801	AAGAAAAAATTAACGATTAAACATT	1825

RESULT	2				
LOCUS	SCU77137				
DEFINITION	SCU77137	1706 bp	DNA	PIN	07-DEC-1996
ACCESSION	Saccharomyces cerevisiae	zinc metallo-protease	(STE24) gene,		
VERSION	U77137	complete cds.			
KEYWORDS	U77137.1	GI:1679740			
SOURCE	baker's yeast.				
ORGANISM	Saccharomyces cerevisiae				

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 1706)	Fujimura-Kamada, K., Nouvet, F. J. and Michaelis, S.	A novel membrane-associated metalloprotease, Ste24p, is required

JOURNAL
REFERENCE
AUTHORS

J. Cell Biol. (1996) In press
2 (bases 1 to 1706)
Michaelis, S. and Fujimura-Kamada, K.

JOURNAL Submitted (01-NOV-1996) Cell Biology and Anatomy, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205, USA

FEATURES	Location/Qualifiers
source	1. .1706

gene	73.1434	.
	/gene="STE24"	
CDS	73.1434	

```

/functional="first step of N-terminal processing of the
yeast a-factor precursor"
/notes="Ste24p"
/codon_start=1
/product="zinc metallo-protease"
/protein_id="AAB38271.1"
/db_xref="GI:16797471"
/translational="MFDKKTLIDHPNIWPKLIISGSFIAQSFESFYSYLYROYOKLSET
KLPLVELEIDDEIDFEFGRSRNYSRAKAFSISGIDVYNLAQKLVFETKYDLPFKIMWAAVS
LNNVLNPFVRHFWSTVAQSLCFGLISLSTPLDPLPSYSHFVLEERGFNGKTLVQD
WITDMIKSLTLAIVGGPIYVFLKIDSPEDDLWYIWEFLVQILAMIIIPVFI
PMENKPTPLEDDELKTKIESIADVGQPKLFIYIDGSKSSHSNAYATGJLPGFSKRI
VLPETLVNSNDSTETVAVLAHEICGQMKQKNNVNNVITLSQHLTELISLFTGISYNTSEF
YNTGGEFLKSTGSGFVDPVYITKEPILIIIGRLPNDLLPLECANQFVMSLSRTHQ
ADAAVAKRLGTRONLCRALIDLIQINKLSTNMNDPLYSTHYSHPLAERLTALDVSEK

```

Query Match	85.1%;	Score 1553.4;	DB 8;	Length 1706;
Best Local Similarity	99.9%;	Pred. No. 2.3e-306;		
Matches 1554;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

```
Oy      271 aaaccactctataatlaacgagaaaggaaaataagaacaattgc   330  
        |||||  
Db       1 AAACCACTTCTATAATTAAcAGGAAAAGAAGAAAAAAGGAAGAATTGAAAACtGC    60
```

Qy	331	aggcccttaacatggtttgtgtcttaagaagatctccaccatcccaataatcccgaggaa	390
Db	61	AGGCGCTTATTCATGTTTGATCTTAAGACGATCTCGACCATCTCAATATCCCGGAA	120
Qy	391	ttactcaatctcgtgtctcgaatctgcocaaattctctcgaactacttgacagaa	450
Db	121	TTAATCATTTCTGGGTTCTCGATTGGCCAAATTTCTTTGGAATCTTACTGACGTACGA	180
Qy	451	cagtaacagaagctatctgaaacaaagtctgcacccctgtctgaaagcgaattgat	510
Db	181	CAGTACCAAGAAGCTATCTGAACAAGATTGCCACCTGTGCTGGAAAGCAAAATGTAT	240
Qy	511	gaaactttcataaataagaagaactactcccgaggccaagggcaagttctcaattcgt	570
Db	241	GAAACTTTTATATAATCAAGAACTACTCCCGGGCCAAAGCCAAAGTTCCATTTTCGT	300
Qy	571	gagctcataaaccagcccaaaagctagttttcataaatacgaactctccctaaatc	630
Db	301	GACGCTATATACCTAGGCCCAAAAGCTATGTTTATATAATAGACCTCTTCCCTAAATC	360
Qy	631	tgagacaatggccgtttctcttatgaaatgcagtcagtcctcagatcataatgtctcc	690
Db	361	TGGCACATGGCGGTTTCTTATTGAATAGCAGTCTGTCGCAATGAGATTTATATATG	420
Qy	691	actctgcacaagattataatgtctctcgtgggtctctatccaagtcttctacacttgat	750
Db	421	ACTGTCCACAGAAATTTATCTTCTTGCGCTCTTATCCAGTTTGCTACCTTTGGTAT	480
Qy	751	ttgcacactcttaccataagcaatttgcctcgtgaagaaatttggtttcataaatgt	810
Db	481	TTGGCACTCTTATCTATACCAATTTTGTCCTGGAAGAAAATTTGGTTTCAATTAATG	540
Qy	811	accgtccaactatgatcaccggaatagatcaagaagctctgacttggcgtatgctatgt	870
Db	541	ACGTTCAACATATGAGATCACCGATATGATCAAGACTGCATTTGGCGATGCTATGTGT	600
Qy	871	ggcccaatccttaccctgtctccttaagaatcttgataaatccactactgattcctctgg	930
Db	601	GGCCCAATCCCTTACCTGTTCTTAAAGATCTTTGAATAATCCCTACATTTCCCTTGG	660
Qy	931	tacatctatgactctctgtctcgttctgtcccaactctagccatgacaaatctccagttc	990
Db	661	TACATTAAGGCTCTGTGTGCTGTGTCACCAATCTTAGCCATGCAATTCACAGTCTTC	720
Qy	991	atcatgcccaatgatttaataagattcaactcacaatctggagacggtgaactgaaaaatctat	1050
Db	721	ATCATGCCCATGTTTATATAGTTACATGCCATTTGGAGAGCGGTGAACGTAAAAATCTAT	780
Qy	1051	gaaagtttggccgataagatgtgggtctccctctagataagaatttctgtaattgaagctca	1110
Db	781	GAAAGTTTGGCCGATAGAGTTGGGTTCCCTCTGTATAGAAATTTTGTCTATGACGGCTCA	840
Qy	1111	aaaagatcttctataacaaacgaatattcaagagtttgcattcaaccctcgaagaatt	1170
Db	841	AAAAGATCTTCTCTTAAACGCAATATTTACAGGTTTGCAATTCACCTCCAAAGAAAT	900
Qy	1171	gtttctgtcagacactttagtgaacagatlaattcaccagatgaaattcagcgttcttgcc	1230
Db	901	GTTTGTTCGACACTTATAGTAAGCAAGATTTACTGATGAATAATTCAGCGCTGTTTGGCC	960
Qy	1231	catagaatcggctcaactgycgaaaaaacaacacacatcgttaataatggtcatctcttagtcaattg	1290
Db	961	CATGAATACTGCTCATCTGGCAAAAAAACACATCTGTAATAATATGCTCTTATAGTCAATTTG	1020
Qy	1291	caagcctctccatcttctccctcttccacagatactaaagaataacatcatctttaaac	1350
Db	1021	CACACCTCTCCCATTTTCTCTCTTTTACCAGCATCTACAGAAATACATCATTTTACAC	1080
Qy	1351	acctctgcgttttctcttaagaagctcaactgycagattctgttgatccggtttcacaatg	1410
Db	1081	ACCTTGCGGCTTTTCTTATAGAAAGTCCACATGGGAGTTTGTGTGATCCGTTATCACTATAG	1140
Qy	1411	gaattcccaatlatcatlgtgattatgtaatttaacgactatlaactccactcgaatgt	1470

/chromosome="1"
/clone="cosmid c3h1"
/map="tl"
complement(join(1..410,455..490))
/gene="SPAC3H1.01c"
/note="or3"
complement(join(1..410,455..490))
/gene="SPAC3H1.01c"
/note="SPAC3H1.01c, len:148"
/codon_start=1
/label=or3
/product="Origin recognition complex-related protein subunit 3."
/protein_id="CA92254.1"
/db_xref="GI:1103502"
/db_xref="SWISS-PROT:Q10067"
/translation="MSA1LQDVSQKTYSVKPKPTESLGNESSEGLFPLIGKEPI EYVLRNLCITKVLNLEIEITHTTYSNARVLRQLEHASPEVDGTLRTAIVCDK SSNSMHLKFEYQLEKDMKEIGKMFVNLINKAHRDLKSCFVIGKS"
complement(411..427)
/note="ctactaattgtag, splice branch and acceptor"
complement(449..454)
/note="gtaaagt, splice donor sequence"
complement(914..4024)
/gene="SPAC3H1.02c"
complement(914..4024)
/gene="SPAC3H1.02c"
/note="SPAC3H1.02c, len:1036, SIMILARITY: Saccharomyces cerevisiae, Y0J8_YEAST, hypothetical 118.4 kd protein, (1037 aa), fasta scores: opt: 1381, E():0, (29.3% identity in 1068 aa)"
/codon_start=1
/label=SPAC3H1.02c
/product="possible zinc metalloproteinase"
/protein_id="CA92255.1"
/db_xref="GI:1103503"
/db_xref="SWISS-PROT:Q10068"
/translation="MLSDTSFROISSEKRLIYVNMKINDATGFSVAVKPTSRLO GSFVATEAHNDNGCPHTLEHLCFMSKRYKPMNGILTFKAGACGDINACDVYTSY ELISAEDGFLRLPVRADHLSPILSDAEACTEYVHINNGEGSSGVYSQMOTQSS ETDVDFCMKRSQPVTSVGYETGHPSELKUSIEKIRYHEKMYPSNICLIVG CINESRLSCASGIVKEILANGIITPTWPRMSTVNDVITPEPILKTVFSEEDV TGSVSLAMNGPSADAVTFAIEFLCELSAVSPGQFVEIEDEPQSFVYVYL RVPSIQLEFSDVPLEKINGLEORALRLAETKOIDNMKMDYKTRRDQVTLNLEF PSSLFMKILITLDHYGSRGKDPNLKMLEYNLLPEMEKDKLKLKKFVNNMY TVIALPSEMAENKKKNAEQLKRLKGLGSGLEAKKSKKENEQKLSNLS SFRISDESIREFYSTARTRSAGQPDNEVQYITDVSESNPFIQDHDSSFVL ATYIDTSMIPSLKPYLSVAKYTEAPALISGTPPYHEVVALQDRDVSILVSLS FDTSSCGAYEYERELLSIEIKYTRNENYGVWIRNLAKTVWDRMLYSINQOL ADIPEKRDAEFLIPSEYDRLYNDGSLKSLANTSOEKLRLRDLKONDSIFDA FOKRMNTMHOAIRIVIGDILKLPQISTWKNLIDSECRNEMPELTFEKNYIK POEGRSSSLTVIWPSPNESSDLVFSIPGTTSWDPELPYVILANTLGLMDGFRNAT IRSGGLAYGEMSIDVGGVLYICINSPDVYKAWASSRDVLEKLGISEGVSSFDLE SAKCVSYISVELENNAIYSSKNSFTLLISGLKKDDEKLEVKQVTLQDFLEGK IYCLPFFSSNNILAVITSSLAKEIQVEQEPFEEGFNVNVEISLHEIQGIESSEDLVY GGNDDN"
complement(3539..3928)
/gene="SPAC3H1.02c"
/note="Match to PF00675 Peptidase_M16, Insulinase (Peptidase family M16) score 21.57"
4683..5123
/gene="SPAC3H1.03"
4683..5123
/gene="SPAC3H1.03"
/note="SPAC3H1.03, len:146"
/codon_start=1
/label=SPAC3H1.03
/product="hypothetical protein"
/protein_id="CA92256.1"
/db_xref="GI:1103504"
/db_xref="SWISS-PROT:Q10069"
/translation="MSLVAYDSEEDQSTLVNENNDIKGRSEEPHAKIPNSPKAEVDT ELEKKIKQFIKLKAKGIHFHTRLSENENFNPRLKLDNIQDFLDIKPRGFMISKMDMD

PTDFHKNVVASALSKSODEMIARRENYOKHDRTEIKFQTSQHPK"
complement(join(5402..5560,5662..7308))
/gene="SPAC3H1.04c"
complement(join(5402..5560,5662..7308))
/gene="SPAC3H1.04c"
/note="SPAC3H1.04c, len:601, SIMILARITY: Candida albicans, O94022, hypothetical 69.0 kd protein., (595 aa), fasta scores: opt: 1014, E():0, (40.0% identity in 598 aa)"
/codon_start=1
/label=SPAC3H1.04c
/product="hypothetical protein"
/protein_id="CA92257.1"
/db_xref="GI:1103505"
/db_xref="SWISS-PROT:Q10070"
/translation="MESRLAKNIARIVQARPKPYLSYHVRASNNALLKOTVIOSS FKSEPTLPRLANRISNSGLISRTTPIITIKQISVRSYSSGDENPEITLKNESNOS GVKRAMPFRVALKMKSLFLFQNKRPVLDVNDVAFESMWLVSHIVATGTTFFSLLY TLNTVSAQELLGRWIGQMLTKNTGFQFSAIYPMNRKGLIFENKLSIARPDPLNG IGQNPNNKSDPEKEKYMALRRKYDSNEPTEALSQNGYQFELSIDKADVSFAF LNKGIYKELOLGRGVVDRRIEMDPSDDPVRKHHMGDFEIKFKLEDRLVTL LQPKGFEPVSVFPCFLPRRKQMLFYDLMNAKTLISFSDNSMFTIHLRLQSPFY LKNGKQIDDMRHSRLRDNDVADILNNGVSGARQINDGSDVLEINISFSPSENF OKAWISLMDKLKKKEDEDYKDVHVNQALHPRKAVIFITNQVSYINNALIRPIT AYINSTRFTFIPILCHVSKPLSDFDGSWTFYDSGVLOEISAQVESEFARDVLANOEIRK RIQKVGYSILRRPLHLVLSLOELAPTTSSISNE"
complement(5561..5576)
/note="ctactttgtagtag, splice branch and acceptor"
complement(55656..5661)
/note="gtaaagt, splice donor sequence"
join(8226..8330,8401..9720)
/gene="SPAC3H1.05"
join(8226..8330,8401..9720)
/gene="SPAC3H1.05"
/note="SPAC3H1.05, len:474, SIMILARITY: Saccharomyces cerevisiae, ST24_YEAST, caax prenyl protease 1, (453 aa), fasta scores: opt: 896, E():0, (40.5% identity in 454 aa)"
/codon_start=1
/label=SPAC3H1.05
/product="putative caax prenyl protease 1"
/protein_id="CA92258.1"
/db_xref="GI:1103506"
/db_xref="SWISS-PROT:Q10071"
/translation="MSPGLCFKIEISYIQATPKPTTSFANCKMGLIQLHMLHIDP GFPMKVIYAGSICKYAMDLYLRROVYLLREKSPALIAEHVEKRYOKALSYARQ SMESTIYSTFLVADLLIKYDGLSYLMTIKTPMMKIKLAASSRSRLTSISPCVF MGLTLESRLIQIPFNLTSTVIEIEKGYSTIKTYITDLKLSLGGGLSMYVAV FVKILTLFGDNFIYAWGAVIVFELLQITAPSLIMPLFKFPLENGSRLTQIEELA ASINPEKTLIVIDASRSRSHNAFFGLPWNKGVLEFDLVKNNHTPELIALIGHEL GHWMSHNLINTIIDGMSLFLPLFAFIRNNLSYSPNITKRPVIVGLILFSDGL GPLSSITLTPASNKYSRLCEYQADAFQAGVAKDLGGILRIHDDNLSLPEFDSLYTS YVHSHPLTVRLNAIDVTTLLKKN"
8331..8336
/note="gtatatt, splice donor sequence"
8386..8400
/note="ctaacgttttagtag, splice branch and acceptor"
8647..8691
/gene="SPAC3H1.05"
/note="PS00678 Beta-transducin family Trp-asp repeats signature"
9106..9690
/gene="SPAC3H1.05"
/note="Match to PF01435 Peptidase_M48, peptidase family M48 score 388.94"
9280..9309
/gene="SPAC3H1.05"
/note="PS00142 Neutral zinc metalloproteinases, zinc-binding region signature"
complement(10343..12112)
/gene="SPAC3H1.06c"
complement(10343..12112)
/gene="SPAC3H1.06c"
/note="SPAC3H1.06c, len:589, SIMILARITY: Schizosaccharomyces pombe, 042922, hypothetical

65.3 kd protein c16a3.17c in chromosome 11., (599 aa),
 fasta scores: opt: 1624, E(): 0, (42.8% identity in 587
 aa)
 /codon_start=1
 /label=SPAC3H1.06c

10.2%: score 187; DB 8; Length 36920;

Best Local Similarity 48.4%: Pred. No. 2,8e-28;
 Matches 664; Conservative 0; Mismatches 655; Indels 54; Gaps 3;

```

OY 340 ttcacgtcttgatccttaagcagatctcgcaccatccctaataatccgylgaatlaacatc 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8392 TTTTATGACATTTGATGACATTTTACGATTTTACGAGCTTCCCTTGAATAATGAT 8451
OY 400 tctgggtctcgtatgcccaatttcttcgaatcttacttactgacgtacagaagacag 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8452 GCTGGAATTTTCTATTGAAAGATATGCTGGGATTTGATCTTGTGTCGACGAGCAATG 8511
OY 460 aagcatctgaaacaaagttgcaccctgctgcgaagacgaatcgaatgaatgaacttt 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8512 TATTTGTTGAGAGAAAGCTCTCTATTTTACGTAACAGCTTGATGATGATGATGATG 8571
OY 520 cataaacaagaactactcccgagcaagcgaagcttcacattcgglyagcgtcat 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8572 CAAAAAGCATTTGCTTATGCTGATGATGATGATGATGATGATGATGATGATGATG 8631
OY 580 aaccagagcccaaaagctagtttccaaatagaccccttcccaaaatccgacatg 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8632 ACTTATGCGCTTGATCTTTGATTTATTAATATGATGATGATGATGATGATGATGATG 8691
OY 640 gccgttcttattga-----atgcagtcctgcagtcacatcatalatgctcc 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8692 ACTAGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8751
OY 691 actgcgcagcagatctatgctctcgtgctcctatccagtttgcacacttgat 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8752 TCGATTAGCATTTCTGTTGTTTCAATGTTGACTTACTCTGTTTCTGTTTCAATG 8811
OY 751 ttcgacactcttactataagccatttgcgcgaagaagaatctggtttcaataatg 810
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8812 ATTCTTTCAATCTTACTCAACCTTTGTGATCGAGGAAAGTATGTTTATTAATACG 8871
OY 811 accgcacaactatgatacagataatgaagaagctgcacttgcgtatgctatg 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8872 ACTTAAAGATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 8931
OY 871 ggcacaatccttaccgttctcctaagatccttgaataatcccaactgattccttgg 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8932 AGTGTGTCGTTGTTGTTTCTGTTTCAATGTTGACCAAGTTTGGTATTAATTTTATG 8991
OY 931 tacattatgctctcgttctcgttgcacaatcttagcatalgacatcactgcttc 990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8992 TATGATGAGGGTCTATATTTTGTGGCTCATTTTACAACATATGCTTCTCTCTA 9051
OY 991 atcatgcccagatgtaataagttcacccatctgagcagcgtgtaacgaataatcat 1050
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9052 ATTATGCCCCCTGTTTACAAGTTTACACCGCTTGAAGAAATGAGAGCTTGAACACA 9111
OY 1051 gaaagtgtggccgaataggttgggtccctcctaagataagatttgcattgagcgcca 1110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9112 GAAAGATTAGCTGCAGATATCAATTTTCTTGAAGGCTTATGATATGATATCTAGT 9171
OY 1111 aaaaatcttcatcacaacacatattcacaaggttgcacatccacccaagaagaat 1170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9172 CGTATATCACTCACTCAACCTCTTTTATATGGCTTTCATGSA---ACAAAGGAAT 9228
OY 1171 gtttgcgcagactttagaagcgaatcttactatgtaaatcagcgttcttggcc 1230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9229 GTACTTTTGAATCTTGTGAAATAATCATAGCAACCGGAACCTTATGCAATCTTGGT 9288
OY 1231 catgaatcgtctactgcgaataaacaacacatcgttataatggtcatttgaatctg 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9289 CACGAACCTTGTCATTGGTACATGTCATTAATTAATTAATTAATTAATTAATTAATG 9348
  
```

```

OY 1291 cacacttccatatttcttcccttccaccgacatcagacgaataacatcatttacaac 1350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9349 ATGAGTCTGTTCACCTCTTCTTATTTGCTGCTTTATTTGAAATAATAGCTTATACACT 9408
OY 1351 accctggccttcttctagagaagtcacatgccaagtttcttgatcccgatcacaaag 1410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9409 TCCTT-----CAATTTCAATTAC 9426
OY 1411 gaattcccatatcatatgattatgattttaaagcattataactccacatgcaatg 1470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9427 GAAAAACCTGTAATTTGTTGACTTCTCTTTTGGACGAGATTTAGTCTTATCTTCC 9486
OY 1471 gccatgcattcgtatgatttatttccagaactcctatgatatcagaagcgtatc 1530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9487 ATCTTACGTTTGTCTCCACAAAGTTTCTGCTTTTGTATGATGATGATGATGATGATG 9546
OY 1531 gctaaaaatgggctacaaacaaatctatgtaggctgaatgattatcctacaatc 1590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9547 GCTAACAATTAAGGTAGCAAGAGATTTGGTGATGTTTAAATAGCATTCATGATGAC 9606
OY 1591 aaccttccacatgatatgatactctgatttctatgatacattatccatccaaact 1650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9607 AATCTTCTCTTGGATTTGATCTTATATACCTGTTATACATGATGATGATGATGATG 9666
OY 1651 ctactgaaagatgcacgcctcctagatatttgtagaagaagaagaacta 1703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9667 TTGGTGTATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 9719
  
```

RESULT 6
 AF353722
 LOCUS
 DEFINITION Arabidopsis thaliana caax processing zinc-metallo endoprotease
 (AF353722)
 ACCESSION AF353722
 VERSION AF353722
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1275)
 AUTHORS
 TITLE
 JOURNAL
 Submitted (27-FEB-2001) Plant Sciences, Tel Aviv University, Ramat
 Aviv, Tel Aviv 69972, Israel
 location/Organism

FEATURES
 source
 1..1275
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="4"
 /map="between markers m122 and g3843; BAC IG002N01"
 1..1275
 /gene="AF353722"
 1..1275
 /gene="AF353722"
 /codon_start=1
 /product="Caax processing zinc-metallo endoprotease"
 /protein_id="AAK39514.1"
 /db_xref="GI:13785909"
 /translation="MAIPMETVGVEMIVIFETVLDLROLAKLPILPKLVGYI
 SOEFKRSRAVSDSYFHVHVEVITLMSALIFRILPRFMSGSAVLPRLGDPE
 NEILHTLSPLAGVWMSQITDLPESLSTVIESRHFQNKQITIMEITRDKGFTLSV
 ILGPVIAAIIETVOKGPLYLALVMEFLLSVMMITVPLVLAFLPKFTPLPDD
 LREKIEKLASLKLPLKLFVVDGSTRSSSNMYMGFFNKRILVLDYDLIOCKND
 EIVAVIAHELGKMLNHTTYSFIVAVIOJLAEFGQGYTLVNRSTDLFSPFEDQPVLI
 GLIIFQHTVPIQHLVSRGINTVSRAPFQADPAFVAVLGAOKLRALVYLQENLSA
 MNDPILSAHTHSRPLVERLRALIDGDKRTD"

BASE COUNT 331 a 266 c 268 g 410 t

Oy	1150	ccattcactccaagaagaatt	1169	
Db	21	CCGTTTCTCTTAACGGAT	2	
RESULT	8			
LOCUS	AR155763/c			
DEFINITION	AR155763	686 bp	DNA	PAT
ACCESSION	AR155763	Sequence 929 from patent US 6239264.		08-AUG-2001
VERSION	AR155763.1	GI:1512816		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 686)			
TITLE	Phillippsen, P., Pohlmann, R., Steiner-Lange, S., Mohr, C., Wendland, J.,			
JOURNAL	Knechtle, P. and Reibischung, C.			
FEATURES	Genomic DNA sequences of ashbya gossypii and uses thereof			
source	Patent: US 6239264-A 929 29-MAY-2001;			
	Location/Qualifiers			
	1..686			
	/organism="unknown"			
BASE COUNT	179 a	173 c	169 g	165 t
ORIGIN				

Query Match	7.5%	Score 136:	DB 6:	Length 666:
Best Local Similarity	52.1%	Pred. No. 7.8e-18:		
Matches 354:	Conservative 0:	Mismatches 320:	Indels 6:	Gaps 2:
Qy 430	ctggaagaacgaatgatgatgaatcatttccataatcaagaactactcccgagccaag	549		
Db 675	CTGGAGGGCGTATTGATTAAGAAACATGCAAAAGACGACGCGTAGACGGGGCCAAAG	616		
Qy 550	gccaaagtctccatllttcgtggaagctataacctagcccaagaagctgatttccataa	609		
Db 615	TAAACGTCACCGTATGGTTAGAGATTCGTGTTCCCGGAGCTAAACTGGTATGATTTAAG	556		
Qy 610	taagaccccttcocccaatctgacacatgagccgtttcttcttaatgaatgacgtccgtca	669		
Db 555	TACAGCTCGGCTGCCACGAGATGTGGAACTAGGAGTGGCGGTAGTGTCAGCG---CATGCC	499		
Qy 670	gtcagaattcatalgtgtcaccatgctgcacagaagttatgcttctgtgtctctatcc	729		
Db 498	GCCATGCTGGTGCCAGTATCGACACTCTCGACGTGGCTGTACTTTTGATGCTACGTA	439		
Qy 730	agttgtctactcgtgtgtgatttgcgaactcttactataagccatttgtctcgtgaaga	789		
Db 438	CAGCTGAACGTGGTGACAGGGGTGTTGGAGACTATTCATCAACTTTGTGCTCAGGAA	379		
Qy 790	aaatttgtttccaataaattgacccgtcccaatgatgatgacagatatcaagaagctg	849		
Db 378	AAATTCCGGGTTCAACAACTCCAGGTCAGTACGTATGCTTACGCGACACAGCTAAAGTGT	319		
Qy 850	aacttgcgtatgcctatgtgtgcccacactccttaacctgttctcctaagaactttgataa	909		
Db 318	ATGATCTCTCCATGATATACCACTCTCCAGCTCTATCGCGTGTGAAGTATTTGAAGAAG	259		
Qy 910	ttccctactgtttcccttctgttacctatgttcttctgttctgttgcgttccaaactttagc	969		
Db 258	TTTCGATGTGGCTTGCTAGCTACGTCCTATACCGTAGCTGTTCTTATCTTGGTGTGG	199		
Qy 970	atgacaacatcttcacgtcttcacacatgcccagtgtttaaataagttacccatctgaagac	1029		
Db 198	ACTGCTCTTCAGCCAGTT---TTACAGCTTTATTTCACCAAGCTACCCCGTTGGAAAC	142		
Qy 1030	gggtgaactgaaaaaactcatgtgaagtttgcgcataagagtgttggttccctatagataag	1089		
Db 141	GGGAGAGCTGAGAGCGTGAATGTGCAATTAAGCAAAACGAGTCAACTTCCACTGGACAA	82		
Qy 1090	attttgtcaattgacgtgtcaaaaagatcttccatcaatcaacgcataatttcaacagtttg	1149		

Db	81	ATCTACCTTTGGGACGGCTCGGCGCCGCTCTGGCGCACTCAATGCGTAATTACCGGGCTA	22
QY	1150	ccattcacctcccaagagaat	1169
Db	21	CCGTTTTCCTCAACGCAT	2
RESULT	9		
E66288/c			
LOCUS		E66288	686 bp DNA
DEFINITION		Genome DNA sequence of Ashbya gossypii and use thereof.	PAT 07-FEB-2001
ACCESSION		E66288	
VERSION		E66288.1	GI:13020629
KEYWORDS		JP 1999225770-A/928.	
SOURCE		unidentified.	
ORGANISM		unidentified.	
REFERENCE		1 (bases 1 to 686)	
AUTHORS		Peter,P.R.P., Steiner,C.M., Juergen,W.P.K.K. and Reblshun.	
TITLE		Genome DNA sequence of Ashbya gossypii and use thereof	
JOURNAL		Patent: JP 1999225770-A 928 24-AUG-1999;	
COMMENT		NOVARTIS AG	
		OS PAG1577RP	

COMMENT	OS	JP	PAG1577RP
	PN	JP 19999225770-A/928	
	PD	24-AUG-1999	
	PF	05-JAN-1998 JP 1998076818	
	PR	31-DEC-1996 CH	16/97
	PI	PETER PHILIPSEN, REINER POHLMANN, SABINE STEINER, CHRISTINE MORE, PI JERGEN WENDLAND, PHILIP KUNDHHTORU, CORINE REBISHUN PC	
		CL12N15/09, CL12R1:865), (CL12N1/15, CL12O1/16, CL12O1/68// (CL12N15/09, CL12R1:645), PC	
		(CL12N15/09, CL12R1:865), (CL12N1/15, CL12R1:645), CL12N15/00, PC	
		(CL12N15/00, CL12R1:645),	
		PC (CL12N15/00, CL12R1:865)	
		CC Strandedness: Single;	
		CC Topology: Linear;	
	FT	Key	Location/Qualifiers
	FT	source	1.686 /Organism='PAG1577RP',
FEATURES			
source		Location/Qualifiers	
		1.686	/organism="unidentified"
		/db_xref="taxon:32644"	
BASE COUNT	179 a	173 c	169 g 165 t
ORIGIN			
Query Match	7.5%	Score 136;	DB 6; Length 686;
Best Local Similarity	52.1%	Pred. No. 7.8e-18;	
Matches 354; Conservative	0;	Mismatches 320;	Indels 6; Gaps 2;
QY	490	ctggagagcaaatgatatgataaactttcataaatcaagaaactactccgggccaag	549
Db	675	ctggagggcgattatgataaagaaacatgcamaagacgagcggtacgagggccacag	616
QY	550	gccaaagtctccatcttcgctgacgctctataacacagcccaaaagctagtttcatcaaa	609
Db	615	taacgctaccctatggtatgagatctgctgtctctgggacataaattcgttaattatgtaag	556
QY	610	tagcaacctctccataaatctgcaaatgacgcttctctattgaaatgaacgactctgca	669
Db	555	tacacatcgctggccacgagatgtgaaccttagagatggcggatgacacgcg--catgcc	499
QY	670	gtcagattcatatcgtctccacgtccgaagaagttatgctctcttggtgctctatcc	729
Db	498	gccattgctggggccagatgacacatctcgagagtcgctgactttttgatgctctactta	439
QY	730	agttgctacactggttattgattgacactctctactataagcaatttgcctcgtgaaagaa	789
Db	438	cagctgaactgctggcagaggggtgtttggagcattattacgtacacatttggctcgagaa	379
QY	790	aaatttgcttcaataaatgacccgtccaactatgatacgcagatagatcaagaagtctg	849
Db	378	aaatttgcttcaataaatgacccgtccaactatgatacgcagatagatcaagaagtctg	319

QY	850	actttggcgtaagcattggtggtgcccacatcccttaacccgttcccttaagaattcttgtaaa	909
Db	318	ATGATCTCCCTCCATATTTAACCACTCTCGACGCTATGCGCTGTGGAAGTAATTGAAGA	259
QY	910	ttccctactgaatttcctttggtacattatgctctcttcttgcgttggtgccaactctagcc	969
Db	258	TTTCGTCAGCTGGCTTGCTGAGCTACGTCCTACCTGATCTGCTGTTCTTCATCTGGGCTG	199
QY	970	atgacaatcatctcagctcttcattcatcgtgccatggttaataagtttaactcattgaggac	1029
Db	198	ACTGCTCTTCAGCCCGTTT--TTTCAGCTTTATTTCACAACTCACCCCGTTGGAGAAGC	142
QY	1030	ggtgacacgaaanaactcatgtgaagtttggtgcccgaataggttggttccctctagtaag	1089
Db	141	GGGAGCTGAAAGACTCGATTGTGCAAAATAGCAAAACGTAACCTCCCACTGGACAA	82
QY	1090	attttgcatgtaagcgctcaaaagaatctctcatcaacgaatattcaacgaatttg	1149
Db	81	ATTCATCTTTGGAGAGGCTCGCGCGCTGTGGGCACTCAATGCGATTTTCAACGGGCTA	22
QY	1150	ccattaccctccaagagaat	1169
Db	21	CCGTTTTCTCTAAACGGAT	2
RESULT	10		
LOCUS	SCYR118C	935 bp	DNA
DEFINITION	S. cerevisiae chromosome X reading frame ORF YR118C.		11-AUG-1997
ACCESSION	Z49618 Y13136		
VERSION	Z49618.1 GI:1015838		
KEYWORDS	baker's yeast.		
SOURCE	Saccharomyces cerevisiae		
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
REFERENCE	1 (bases 1 to 935) Rose, M., Koeltter, P. and Entlian, K. D. Unpublished		
AUTHORS	2 (bases 1 to 935) MIPS.		
JOURNAL	Direct Submission		
TITLE	Submitted (25-SEP-1995) Data collected by MIPS on behalf of the		
REFERENCE	European yeast chromosome X sequencing project. MIPS at the		
AUTHORS	Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a D-82152		
JOURNAL	Martinsried, FRG; E-mail: Mewes@mips.emblnet.org		
FEATURES	Location/Qualifiers		
SOURCE	1..935		
CDS	/organism="Saccharomyces cerevisiae"		
	/db_xref="taxon:4932"		
	/chromosome="X"		
	complement(122..733)		
	/note="ORF YR118C"		
	/codon_start=1		
	/protein_id="CA89664.1"		
	/db_xref="GI:1015839"		
	/db_xref="SWISS-PROT:P47155"		
	/translation="MAQALNSTINIAFFVAFLETTIAFCLKNVNSILONTFTIVH		
	MLEQLTSLRYSGQLGFALFTLNGVDLPLLENNYKYOSVVPVRLTFEII		
	SLYSESNFYVNNNSVFTYCFAEVWINFLINLAIREKKEEFRRLNQFVNDEIDIE		
	PFYKTEETLEIILINDENDDEGKDDNDNNEKGNDSDAK"		
BASE COUNT	316 a	159 c	138 g
ORIGIN			322 t

Query Match	6.6%	Score 121	DB 8	Length 935
Best Local Similarity	100.0%	Pred. No.	8/7e-15	
Matches 121	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy 1705	tctatagagtcacactattatgacatgctacgcgttaaattcagcttcgttatgctatctata	1764		
Db 1	TCATATAGAGTCACACTATTATGACATGCTTAAATTCAGCTTCGTTATGCTATATCTCA	60		

QY	1765	cacatacaacaggtatcctactctaagaatlaaaaggaaagaaaanaataaacgattlaaacat	1822
Db	61	CATCATACACAGGTATCTCTAATAAGAATAAAAGAAAATAATAACGATTAAACAT	120
QY	1825 ± 1825		
Db	121 T 121		

RESULT	11
LOCUS	AF064867
DEFINITION	AF064867 Homo sapiens CAXX prenyl protease (STE24) mRNA, complete cds.
ACCESSION	AF064867
VERSION	AF064867.1 GI:3800768
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1859) Tam,A., Nouvet,F.T., Fujimura-Kamada,K., Siunt,H., Sisodia,S.S. and Michaelis,S. Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAXX processing J. Cell Biol. 142 (3), 635-649 (1998) 963653461 2 (bases 1 to 1859) Tam,A., Nouvet,F.T., Fujimura-Kamada,K., Siunt,H., Sisodia,S.S. and Michaelis,S. Direct Submission Submitted (12-MAY-1998) Cell Biology and Anatomy, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205, USA Location/Qualifiers
FEATURES	

	Query Match	Best Local Similarity	Matches 370; Conservative	Score 117.4; DB 9; Length 1859; Pred. No. 4.6e-14; Mismatches 351; Indels 12; Gaps 3;
Qy	471 aacaaattgacactgtctcgtgaagacgaatlgatgatataaacttttcaataatcaag	50.5%;	0;	530
Db	267 AACTCATGTACCACCGGAGTTAGGACAGATCATGATGATTCGAAACATTTGGAATAATTCG			326
Qy	531 gaactactcccgagcgcaagcgcaagtttcattcatttcggtgagctctataactcagccca			590
Db	327 ACTGTATCACTGTGATAAAGACCATCTTACGTTTCGGTCAGACACTCTATTTCAGAGACTGA			386

Db 913 TGACTTTCCTTGACGAGGTGTATGTGTGAGAGACTCTAAGCTTTCACAGCAA 972
 QY 1131 CGCATATTCACAGGTTTGCATCACCACAGAGAAATGTTGTTGACACCTTACT 1190
 Db 973 TCCATTATTTTATGGCTT---CTTCAGAGCAAGCAAGATGTTTGTGACACCTACT 1029
 QY 1191 gaacagtaattct 1203
 Db 1030 AGAAGAGTACTCT 1042.

RESULT 13

HSY13834 2966 bp mRNA PRI 01-JUL-1999
 LOCUS HSY13834
 DEFINITION Homo sapiens mRNA for farnesylated-proteins converting enzyme 1.
 ACCESSION Y13834 GI:5327058
 VERSION 1
 KEYWORDS farnesylated-proteins converting enzyme; gluzincin; metalloproteinase.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2966)
 AUTHORS Freije,J.M., Blay,P., Pendas,A.M., Cadianos,J., Crespo,P. and L
 pez-Ortin,C.
 TITLE Identification and chromosomal location of two human genes encoding
 enzymes potentially involved in proteolytic maturation of
 farnesylated proteins
 JOURNAL Genomics 58 (3), 270-280 (1999)
 MEDLINE 99303558
 REFERENCE 2 (bases 1 to 2966)
 AUTHORS Lopez-Ortin,C.
 TITLE Direct Submission
 JOURNAL Submitted (13-JUN-1997) C. Lopez-Ortin, Universidad de Oviedo, Dept.
 de Bioquímica y Biología Molecular, Facultad de Medicina, C/ Julian
 Claveria S/N, 33006-Oviedo, SPAIN

FEATURES

1..2966
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="LAMBDA DR2"
 /clone="A-1.3a"
 /tissue_type="Ovary"
 34..1461
 /gene="facel"
 34..1461
 /gene="facel"
 /codon_start=1
 /product="farnesylated-proteins converting enzyme 1"
 /protein_id="CAB46277.1"
 /db_xref="GI:5327059"
 /translation="MGMAASLDALMEMPAERKIFGAVLLFSWTYVLMETFLAQRORI
 YKTTTHVPELGQIMDSPEKSLRYQDKSTFSWSGLYSEETGLILLGGLPIYL
 RLSGRCYAGEPYETIOSLVFLATLPLMSLYNTPVIEKGFNOOTL
 GPFMKDAIKKFFVOCILLPVLSILYIKIGDGFYAMLFITVIEGLVITADY
 IAPLEKRPPLPEKLEIEYMAKSIDPPLKRYVVGSKRSSNAKPTGFFRNKR
 IVPDLLEETSVLNKIDQEDSGMEPRNEEGNSEIYAKVKNKQCKNEVLAVIG
 HELGHWKLGHTVKNIISQMSFLCFPLFAVLIGKRELFAAGFGLDPTLGLIIF
 OFISPYNEVLSCITVLSRREFQADAFKAKGLKAKDLYSALLIKLNKDNLGFPVSDM
 LFSMWHYHPLELEOLAKTMKOH"
 BASE COUNT 844 a 515 c 578 g 1029 t
 ORIGIN

Query Match 6.4%; Score 117.4; DB 9; Length 2966;
 Best Local Similarity 50.5%; Pred. No. 4.5e-14;
 Matches 370; Conservative 0; Mismatches 351; Indels 12; Gaps 3;

QY 471 aacaaagtgcacactgtcgtgaagaaattgatgataaactttcataaataacg 530
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 177 AACCATGATACACCGAGCTAGACAGATCATGATCTCTGAACATTTGAAATAATCTCG 236

QY 531 gaactactccgggccaaggaagcttcacatttcggtgacgtatataactgacca 590
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 237 ACTCTATACAGTGGATTAAGCATCTTACAGCTTGTGTAAGACTATTTACAGACTGA 296
 QY 591 aaagctagttcatcaataacagacctccccaataacatgcacatgacggttctt 650
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 297 AGGACACTTATTTCTCTCTTTGGAGGAAATACCTTATCTCTGAGACTTTTCGAGGTT 356
 QY 651 attgaatgcagtcctgcagcaagatltcatatgltcccaactgltgcagaaagtatg 710
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 357 CT--GTGGTTATCTGGCTTTTGACCCAGATATGAGATCATCTAGTCCCTGGGTTTCG 414
 QY 711 cttcttggtctctatccagattgttcacacctgtgtgtgtgtgtgtgtgtgtgtgt 770
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 415 CTGTTGGCTTACACT-----TTTCAGTGCATTGACTGTTTGCCTGAGAGCTTTATATA 467
 QY 771 ccatlttgctcggagaaataattgtgtcaataaataatgcacgtlccaaactatgatac 830
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 468 TACTTTTGTGATAGAGAAACATGCTTCATATCAGACACTTTGGGTTCTTCATGAA 527
 QY 831 cgatatgatcaagagtcgactttggcgatgctatgtgtggtggtggtggtggtggt 890
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 528 AGATGCAATCAAGAAATTTGTTGACTCAGTATTTTGTGCTGTGCTTCACTTCT 587
 QY 891 ccttaagatccttgataaattccctacatgatttcccttggtgataatggttctgtgt 950
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 588 ACTTTACATTTATTAATTTGGGGGTGACTATTTTATTTATTTAGCTGCTGTTCACAT 647
 QY 951 cgtgtccaaacttgacatcagatcagatcagatcagatcagatcagatcagatcagat 1010
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 648 AGTTGTGCTGCTGCTTCTGTCACATCTATGCTGATTTATTTGCTGCTTATTTGACAA 707
 QY 1011 gttcaactcattggaagagcgtggaactgaaataatcattgaagtgttgcgataag 1070
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 708 ATTCAACCTCTGCCCTGAGGGAAGCTTAAGAAAGAAATGGAAGTAAATGGAAGTAT 767
 QY 1071 tgggtccctcctagataaattttgtcatltagcaggtcgaataaagatccttcataca 1130
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 768 TGACTTTCCTTGACGAGGTGTATGTTGTGAGAGATCTTAACGCTTTCCACAGCAA 827
 QY 1131 CGCATATTCACAGGTTTGCATCACCACAGAGAAATGTTGTTGACACCTTACT 1190
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 828 TCGTATTTTATGGCTT---CTTCAGAGCAAGCAAGATGTTTGTGACACCTACT 884
 QY 1191 gaacagtaattct 1203
 Db 885 AGAAGAGTACTCT 897

RESULT 14

E32056 2968 bp DNA PAT 07-FEB-2001
 LOCUS E32056
 DEFINITION Human AFCL.
 ACCESSION E32056
 VERSION E32056.1 GI:13026656
 KEYWORDS JP 1999113579-A/1.
 SOURCE JP 1999113579-A/1.
 ORGANISM unclassified.
 1 (bases 1 to 2968)
 REFERENCE 1 (bases 1 to 2968)
 AUTHORS Kristlin,K.K.C.C. and Sazan,A.M.N.N.
 TITLE Human AFCL
 JOURNAL Human AFCL
 PATENT: JP 1999113579-A 1 27-APR-1999;
 SMITHKLINE BEECHAM CORP,SMITHKLINE BEECHAM CORP PUBLIC LTD CO
 OS Unidentified
 PN JP 1999113579-A/1
 PD 27-APR-1999
 PF 24-JUN-1997 GB 97304440:7.12-FEB-1998 US 09/022699 PI
 KRISTIN K KIRKLY,CHRISTOPHER D SAZAN,ANN M NABU PC
 C12N15/09,A61K38/00,A61K39/395,A61K48/00,C07K14/47, PC
 C07K16/18,
 PC C12N9/64,C12Q1/68,G01N33/15,G01N33/53,C12N15/00,A61K37/02 CC

Db 816 TTGTCTCTACGACACTTTGATTAACCAAGTCGACACTCCGGCATTTGCGCGCTTAG 875
QY 1229 cccatgaaatcgcgtcactgcaaaaaaacacacgcgttaataatgltcatttagtcaat 1288
Db 876 CCCACGAACTTGGACATTTCAAAATGAGCCACACGTAACAAGAACCTGGTCTCTACGCGAGC 935
QY 1289 tgcacaccttccatcttcccttcccttccacagcaltcaagaatcacatcatttaca 1348
Db 936 TCTACATGCTCGTCTTTTATTCCTTTTCAAGCCAGTCGCTTTTAACGTGATTTATACC 995
QY 1349 acaccttcggccttcttcttagag 1371
Db 996 GGAGTTTCGGATTTCCACCTGAG 1018

Search completed: February 12, 2002, 13:20:11
Job time: 9719 sec

THIS PAGE BLANK (13PTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:32:24 ; Search time 17.02 Seconds

(without alignments)
678.579 Million cell updates/sec

Title: US-09-165-460A-4

Perfect score: 1651

Sequence: 1 MGFSTFLVLYISISYLP.....ISLKDPLQTLVGTGPGRTTL 315

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1651	100.0	315	1	003530 saccharomyc
2	249	15.1	290	1	FAC2_DROME
3	230.5	14.0	329	1	FAC2_MOUSE
4	229.5	13.9	329	1	FAC2_MOUSE
5	202.5	12.3	271	1	RCEL_SCHPO
6	100.5	6.1	433	1	NU4M_RHISA
7	99	6.0	321	1	MRAY_STRAU
8	96.5	5.8	226	1	ATP6_ANOGA
9	96.5	5.8	534	1	Y25 MYCPN
10	96	5.8	331	1	GPR1_HUMAN
11	96	5.8	426	1	YAB3_SCHPO
12	95.5	5.7	224	1	ATP6_DROSI
13	94.5	5.7	224	1	ATP6_DROSI
14	94.5	5.7	312	1	OLP6_CHICK
15	94	5.7	226	1	ATP6_ANOU
16	93	5.6	2376	1	YIM6_YEAST
17	92	5.6	715	1	NU5M_NECR
18	91.5	5.5	471	1	5H2A_HUMAN
19	90.5	5.5	456	1	ENT2_HUMAN
20	90.5	5.5	471	1	5H2A_MACMU
21	89.5	5.4	471	1	5H2A_CRIGR
22	89.5	5.4	471	1	5H2A_MOUSE
23	89.5	5.4	471	1	5H2A_MOUSE
24	89	5.4	347	1	NU2M_DIDMA
25	89	5.4	347	1	NU2M_DIDMA
26	89	5.4	429	1	RNE_GUITH
27	89	5.4	2232	1	CCAE_RAT
28	88.5	5.4	312	1	OLP6_RAT
29	88.5	5.4	318	1	OLP6_CHICK
30	88.5	5.4	433	1	Y681_PASMU
31	88.5	5.4	1882	1	Y681_MYCPN
32	87.5	5.3	821	1	CAN3_RAT
33	87	5.3	364	1	ERG3_CANCA

34	87	5.3	470	1	5H2A_PIG
35	86.5	5.2	334	1	YB77_METJA
36	86.5	5.2	599	1	DNAB_GUITH
37	86.5	5.2	637	1	NU5M_STRPU
38	86.5	5.2	3655	1	YAB3_SCHPO
39	86	5.2	342	1	PAFR_HUMAN
40	85.5	5.2	312	1	OLP6_CHICK
41	85.5	5.2	583	1	HASS_XENLA
42	85.5	5.2	712	1	FREP_YEAST
43	85.5	5.2	821	1	CAN3_MOUSE
44	85.5	5.2	974	1	ATXA_LEIDO
45	85.5	5.2	974	1	ATXB_LEIDO

ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	315 AA
1	RCEL_YEAST	003530			
AC	RCEL_YEAST	003530			
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	CAAX PRENYL PROTEIN 2 (EC 3.4.22.-) (PRENYL PROTEIN-SPECIFIC				
DE	ENDOPROTEASE 2) (PPSEP 2) (RMS AND A-FACTOR CONVERTING ENZYME) (RACE).				
OS	RCEL OR YMR274C OR YMR156.16C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID:4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN:5288C / AB972;				
RA	Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.;				
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	FUNCTION.				
RX	MEDLINE:97218305; PubMed-9065405;				
RA	Boyartchuk V.L., Ashby M.N., Rine J.;				
RT	"Modulation of Ras and a-factor function by carboxyl-terminal				
RL	proteolysis.";				
RL	Science 275:1796-1800(1997).				
RN	[3]				
RP	SUBCELLULAR LOCATION.				
RX	MEDLINE:98409630; PubMed-9736709;				
RA	Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.;				
RT	"Endoplasmic reticulum membrane localization of Rcep1 and Ste24p,				
RL	yeast proteases involved in carboxyl-terminal CAAX protein processing				
RL	and amino-terminal a-factor cleavage.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).				
RN	[4]				
RP	CHARACTERIZATION.				
RX	MEDLINE:20285442; PubMed-10825201;				
RA	Trueblood C.E., Boyartchuk V.L., Picooglou E.A., Rozema D.;				
RT	"The Caax proteases, Afc1p and Rcep1, have overlapping but distinct				
RL	substrate specificities.";				
RL	Mol. Cell. Biol. 20:4381-4392(2000).				
CC	-1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF				
CC	-1- FARNESYLATED A-FACTOR MATING PHEROMONE AND OF RAS.				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC				
CC	RETICULUM.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52.				
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				

DR EMBL: 249260; CAA89257.1; -
 DR MEROPS: C52.001; -
 DR SGD: S0004887; RCEI.
 KW Hydrolase; Thiol protease; Transmembrane; Endoplasmic reticulum;
 KM Phosphomono response.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 75 95 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 238 258 POTENTIAL.
 FT TRANSMEM 276 296 POTENTIAL.
 SO SEQUENCE 315 AA; 35911 MW; D151F4790F06081 CRC64;

Query Match 100.0%; Score 1651; DB 1; Length 315;
 Best Local Similarity 100.0%; Pred. No. 2,7e-123;
 Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLOPSTFLVLLYISIVLPLATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVLPLO 60
 1 MLOPSTFLVLLYISIVLPLATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVLPLO 60
 61 SOLSTTSHSFKDAFLGLITPGYVALPMPQSFVADLTCKVAMLLTYCGPYLDF 120
 61 SOLSTTSHSFKDAFLGLITPGYVALPMPQSFVADLTCKVAMLLTYCGPYLDF 120
 121 VLVHLPKSKSLLEDFEHEFLNIMSFNRTFAPITTEIFTSMLTYLNLPHSOLSYQ 180
 121 VLVHLPKSKSLLEDFEHEFLNIMSFNRTFAPITTEIFTSMLTYLNLPHSOLSYQ 180
 181 QLEWQPSLFFGLAHAAHAYEOLGSGMTYSILTTFCQILYTLFGLTKFVFPRTGCGN 240
 181 QLEWQPSLFFGLAHAAHAYEOLGSGMTYSILTTFCQILYTLFGLTKFVFPRTGCGN 240
 181 QLEWQPSLFFGLAHAAHAYEOLGSGMTYSILTTFCQILYTLFGLTKFVFPRTGCGN 240
 241 LMCCIIHALCNMGFPGPSRLNHTVYDKKAGRIKLVSNMKCYFALLVLGLSLKD 300
 241 LMCCIIHALCNMGFPGPSRLNHTVYDKKAGRIKLVSNMKCYFALLVLGLSLKD 300
 241 LMCCIIHALCNMGFPGPSRLNHTVYDKKAGRIKLVSNMKCYFALLVLGLSLKD 300
 301 TLQTVGTPGRTTL 315
 301 TLQTVGTPGRTTL 315
 301 TLQTVGTPGRTTL 315

RESULT 2
 FAC2_DROME STANDARD; PRT; 290 AA.
 AC 09U1H8; 09VBM4;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE CAAX PREVYL PROTEASE 2 (EC 3.4.22.-) (PREVYL PROTEIN-SPECIFIC
 ENDOROTASE 2) (FARNESYLATED-PROTEIN CONVERTING ENZYME 2) (PAGE-2)
 DE (SEVERAS PROTEIN).
 GN SRAS OR CG4852.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Weizsaeckel-Hinun A., Toerock I., Kiss I., Fairas R., Mechler B.M.;
 RT "The severas gene of Drosophila encodes a CAAX-protease and acts as a
 RT tumour suppressor."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H., Blazee R.G., Champe M., Pfeiffer B.D.,
 RA Wan R.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Adair J.F., Agbayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mitsuhashi N.V., Mobarry C., Morris J., Mosnig A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley R.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED AND GERANYLATED PROTEINS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AJ252068; CAB64383.1; -
 DR EMBL: AE003565; AAP50770.2; -
 DR FlyBase: FBgn0029121; Stras.
 DR InterPro: IPR003675; Abi.
 DR Pfam: PF02517; Abi. 1.
 DR Hydrolase; Thiol protease; Transmembrane; Endoplasmic reticulum.
 FT TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 159 179 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 FT TRANSMEM 256 276 POTENTIAL.
 FT TRANSMEM 82 83 POTENTIAL.
 FT TRANSMEM 110 110 POTENTIAL.
 FT TRANSMEM 126 126 POTENTIAL.
 FT TRANSMEM 126 126 POTENTIAL.
 SO SEQUENCE 290 AA; 33082 MW; 157CAAS5B30693A8 CRC64;

Query Match 15.1%; Score 249; DB 1; Length 290;
 Best Local Similarity 29.1%; Pred. No. 7.8e-13;
 Matches 75; Conservative 52; Mismatches 83; Indels 48; Gaps 14;


```

QY      8  LVLVLTSLSTSVLPYAAVSOPGKSRNDPRTKSKMOKLTINLSNLPVLFQSLSTT  67
Db      25  LAVLVGSLYI--WSTKH-----NRDPTTYKKRFASVSNMMLAAPFVFFSS--PELL  75

QY      68  SHISFKDAFLGGLIPGYAALPNPWOFSQFKDLTKCVAMLLTYGCPVLDPVLYHLN  127
Db      76  SRVPF-PKLLGLR-LEGIMQAVPIPY-----SLTVLLFGPI-----FVN  113

QY      128 PKSSILEDVEYHEFINIM--SF-----RNFAPRTETEEIYTSMLLTYINLPHQSLSY  179
Db      114 MQNSVSKSYF--LDYMRGSGSTIWRNHNHIAFLSEEVFRACMP-----LTIQSFPL  167

QY      180  QQLFMOESFLPGLANHANNAVEQLOEGSMITVYSILLTTCFOILYTTLEGTLTKFEVFTGG  239
Db      168  VAVETTP-LRFGVANHNIHIERLSLGVELSPALLI-GLPQRTYVTLRFGFSAFLLPAT-G  224

QY      240  NLKCCIIILHAIICNMGEP  257
Db      225  HVMAPIILVHAFICNMGRL  242

```

	RESULT	3	
ID	FAC2_MOUSE		PRT: 329 AA.
AC	p57791; Q9PE68;	STANDARD:	
DT	20-AUG-2001 (Rel. 40, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	CAX PRENL PROTEASE 2 (EC 3.4.22.-) (PRENL PROTEIN-SPECIFIC ENOPROTEASE 2) (FARNESYLATED-PROTEINS CONVERTING ENZYME 2) (FACE-2).		
OS	RCE1A OR RCE1 OR FACE2.		
NCBI_TaxID	10090;		
OC	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Cadinanos J., Freije J.M.P.;		
RT	"Characterization and expression analysis of the gene encoding the		
RT	murine Cax protease face-2.";		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	KNOCK-OUT.		
RX	MEDLINE-99185053; PubMed-10085069;		
RA	Kim E., Ambrozik P., Otto J.C., Taylor B., Ashby M., Shannon K.,		
RA	Casey P.J., Young S.G.;		
RT	"Disruption of the mouse rce1 gene results in defective Ras processing		
RT	and mistocalization of Ras within cells.";		
RL	J. Biol. Chem. 274:8383-8390(1999).		
CC	-I- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF FARNESYLATED AND GERANYLATED PROTEINS. SEEMS TO BE ABLE TO PROCESS K-RAS, N-RAS, H-RAS, RAP1B AND G-GAMMA-1 (BY SIMILARITY).		
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM (BY SIMILARITY).		
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-slb.ch/announce/ or send an email to license@isb-slb.ch)		
CC	-----		
DR	EMBL; AJ251644; CAC17013.1; -		
DR	EMBL; AJ251645; CAC17014.1; -		
DR	MEROPS; C52_002; -		
DR	MCD; MG1:1336895; Rcel.		
KW	Hydrolase; Thiol protease; Transmembrane; Endoplasmic reticulum.		
FT	TRANSMEM 36 56 POTENTIAL.		
FT	TRANSMEM 75 95 POTENTIAL.		
FT	TRANSMEM 112 132 POTENTIAL.		
FT	TRANSMEM 193 213 POTENTIAL.		

FT	TRANSMEM	226	246	POTENTIAL.
FT <th>TRANSMEM</th> <td>254 <td>274 <th>POTENTIAL.</th> </td></td>	TRANSMEM	254 <td>274 <th>POTENTIAL.</th> </td>	274 <th>POTENTIAL.</th>	POTENTIAL.
FT <th>TRANSMEM</th> <td>282 <td>302 <th>POTENTIAL.</th> </td></td>	TRANSMEM	282 <td>302 <th>POTENTIAL.</th> </td>	302 <th>POTENTIAL.</th>	POTENTIAL.
SO <td>SEQUENCE</td> <td>329 AA;</td> <td>35866 MM;</td> <td>C2A49617EBC77EC8 CRC64</td>	SEQUENCE	329 AA;	35866 MM;	C2A49617EBC77EC8 CRC64

Query Match	14.08;	Score 230.5;	DB 1;	Length 329;
Best Local Similarity	30.68;	Pred. No. 2.6e-11;		
Matches 85; Conservative	40;	Mismatches 82;	Indels 71;	Gaps 16

QY	4	ESTPLVLLYSIVLEPLVLTATSOEGRSKRNDPRFKRMRQKLTMTLSINF-LVPEFQSQ	62
Db	40	FSCF-----SLACSVGSLVYWKSE--LPRDHRAIKRR--STVLVVSISLPLCYLLMRE	91
QY	63	LSST---TSHISEKDAFLGLGIIIPGYAALRNPMQSFQWKDLTKCYAMLLT-LYCGPV	117
Db	92	LTGIQPTSLTLLM-GRLEGIFIPA--ALLP-----LLTMTLLEGLRP	131
QY	118	LDPV-----LYNLLMPKSSILEDYFNEFLNWS-----FRNEFPAPTEEIF	139
Db	132	MQLSMDCRDLTDLGKAVLAPRS-----MARCLTDKRWLRNOVIAPLDELV	178
QY	160	YTSMLLTYYLNLPHSOLSYQOELFMORSLFFGLAHANHAAYEOLQEGSMTYSILLTTCFQ	219
Db	179	FRACMLPM--LAPCTGIG-PAVETCP-LREGVANHNIIEQLRFRQSSVGSIFVSAFQ	233
QY	220	ILYTLTFEGSLTKFVAVTGGNLMCCITLLAHLCNMGRR	257
Db	234	ESTYAVFGAATALELIFT-GHLIGRPVLCSPSCNMGRP	270

	RESULT	4		
ID	FAC2_HUMAN		PRT;	329 AA.
AC	G9Y2S6;	STANDARD;		
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CXAX PRENYL PROTEASE 2 (EC 3.4.22.-) (PRENL PROTEIN-SPECIFIC			
DE	ENDOPOTEPHASE 2.) (FARNESYLATED-PROTEINS CONVERTING ENZYME 2) (FACE-2).			
GN	RCE1A OR RCE1 OR FACE2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Ovary:			
RX	MEDLINE=99303558; PubMed=10373325;			
RA	Frelle J.M.P., Blay P., Pendas A.M., Caninanos J., Crespo P.,			
RA	Lopez-Olin C.;			
RT	"Identification and chromosomal location of two human genes encoding			
RT	enzymes potentially involved in proteolytic maturation of farnesylated			
RT	proteins."			
RL	Genomics 58:270-280(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99185052; PubMed=10085066;			
RA	Otto J.C., Kim E., Young S.G., Casey P.J.;			
RT	"Cloning and characterization of a mammalian ptenyl protein-specific			
RT	protease."			
RL	J. Biol. Chem. 274:8379-8382(1999).			
CC	-1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF			
CC	FARNESYLATED AND GERANYLATED PROTEINS. SEEMS TO BE ABLE TO PROCESS			
CC	K-RAS, N-RAS, H-RAS, RAP1B AND G-GAMMA-1.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC			
CC	RETICULUM (BY SIMILARITY).			
CC	-1- TISSUE SPECIFICITY: UBIQUITOUS.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y13835; CAB46278.1; -
DR EMBL; AF121951; AAD22632.1; -
DR MEROPS; C52.002; -
DR MIM; 605385; -
KW Hydrolyase: Thiol protease; Transmembrane; Endoplasmic reticulum.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
SQ SEQUENCE 329 AA; 33632 MW; BA8F764651172BFA CRC64;

Query Match 13.9%; Score 229.5; DB 1; Length 329;
Best Local Similarity 30.6%; Pred. No. 3,1e-11;
Matches 83; Conservative 39; Mismatches 82; Indels 67; Gaps 15;

QY 11 LYSISVPLPLVATSPQSGKRDNPRTIKSRMOKLTITLISNLF-LVPFLOSQSLST--- 66
DQ 43 LSLACSYVGSGLYWKSE--LPRDHAVIKRRF--TSVLVSSLSPLCLVLRRLTGIDPG 98
QY 67 TSHISFKDAFLGDIIPGYAALPNPQFSQFVNDLTRCVAMLT--LYCGPVLDV-- 121
DQ 99 TSLTLTLG-GEFLGIFPA--ALLP-----LLTMTLIFGLPLQQLSMDC 138
QY 122 -----LYHLNPKSSLTLEDYHHEFLNINS-----FNFETFAPIETEEFTSMILT 166
DQ 139 PCDLADGKAVVLARS-----WAKCLDMKMLKNQVIAPILEELVRRACMLP 185
QY 167 TYLNLIPHSQSLSYOQLFQWPSLFFGLAHAAHAYEQLQSGMTVSILTTFCQLLYTLF 226
DQ 186 M---LAPCMGLG-PAVFICP-LFGVAHFHIIIDQLRRSSVGNIFLSAASFQSYTAVF 240
QY 227 GGLTFVVRGTGNNMCCIIHALCNIMGFP 257
DQ 241 GAYTAFLEIRT-GHLIGPVLCHECNYMGFP 270

RESULT 5
RCBL_SCHPO STANDARD; PRT; 271 AA.
AC 094448;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GAAK PRENYL PROTEASE 2 (EC 3.4.22.-) (PRENYL PROTEIN-SPECIFIC
DE ENDOPROTEASE 2) (PPSEP 2).
GN SPAC1687.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Medlert H., Mambult R., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
CC FARNSYLATED PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL035064; CAA22596.1; -
KW Hypothetical protein; Hydrolyase; Thiol protease; Transmembrane;
KW Endoplasmic reticulum.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
SQ SEQUENCE 271 AA; 31024 MW; 7274E9EE223448BB CRC64;

Query Match 12.3%; Score 202.5; DB 1; Length 271;
Best Local Similarity 23.9%; Pred. No. 3,4e-09;
Matches 74; Conservative 55; Mismatches 94; Indels 87; Gaps 14;

QY 7 FLVLLYISVPLPLVLA--TSQEGS--KRDNPRTIKSRMOKLTITLISN-----LFLV 56
DQ 4 YLISFFFAIVVSLYFFPVAPRPSLNRPNDKVIYAKC--ISVLLASSVCCILTRLIG 61
QY 57 PFIQSQSLSTSHISFKDAFLGDIIPGYAALPNPQF--SQFVKDLTKCVAMLTLYC 114
DQ 62 PSL-----NVTFFPIPDYVLSKLHNA-----TIFI 86
QY 115 GPVLDFVLYHLNPKSSLTLEDYHHEFL-----NWSFNFETFAPIETEEFTSMILT 166
DQ 87 GP-----LYEYV-----IVDKERYLFTHLKDCLSNIAWRNIILGISELTERCCIVP 136
QY 167 TYLNLIPHSQSLSYOQLFQWPSLFFGLAHAAHAYEQLQSGMTVSILTTFCQLLYTLF 226
DQ 137 -----TCENAGSRILKTIIFVADPLFGMAHHTHTYFLLAYPNAYIAALQYVQPSYTVF 192
QY 227 GGLTFVVRGTGNNMCCIIHALCNIMGFSPRLNHFTVYDKKGRISKLSIWNKC 286
DQ 193 GWYTHLFLST-HSLFSPFLVHACNSMGLP-----TLXGIGRNRQTRI----- 236
QY 287 YEALLVGLI 296
DQ 237 YVTLILGLV 246

RESULT 6
ND4M_RHISA STANDARD; PRT; 433 AA.
AC 099825;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3).
GN ND4.
OS Rhinipcephalus saungineus (Brown dog tick).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Rhinipcephalus.
OX NCBI_TaxID=34632;
RN [1]
RP SEQUENCE FROM N.A.
RA Black W.C. IV, Roehrdanz R.L.;
RT "Mitochondrial gene order is not conserved in arthropods: prostriate
RT and metastriate tick mitochondrial genomes.";
RL Mol. Biol. Evol. 15:1772-1785(1998).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

[illegible]

Db 70 VDO--ANDPILLFVITGIFGIDDIYK--KNOGLTSKQELAOI-----GIA 119
 QY 117 VLDPVL---YHLNPKSSILEDPEHEFLNI---MSFRNFAPITEEIFYTSMLTTYLN 170
 Db 120 IIFVLSNVFLHVNFSI-----HIPFNVAIPISFAVIFI-----157
 QY 171 LIPHSLSYQOLFQPSLFFGLAHNAHAYEQLQBSMTTVSILLTTCFOILYTLFGGLT 230
 Db 158 -----VFMO-----VGFSSNAVNLTDGL-DGLATGLSIIIGFTYMAIM-----192
 QY 231 KEFVVRGNGNIMCCIIILHALCNMGFP-GPSRLNHFVYVKKAGRISKVSTNNKCYFA 289
 Db 193 SFVAGETAIGFICIMFALLGFLPYNNIPAKVEMGDT-GSLAGLGFATISIMLNDELS 251
 QY 290 LVLGLISLKDITLQTVCTPCGRIT 314
 Db 252 LIFGLVFEVITELSVMLQVASFRLT 276

RESULT 8

ID ATP6_ANOGA STANDARD; PRT; 226 AA.
 AC P34834;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
 GN ATP6
 OS Anopheles gambiae (African malaria mosquito).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 OC Anopheles.
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G3;
 RX MEDLINE=97242550; PubMed=9087549;
 RA Beard C.B., Hamm D.M., Collins F.H.;
 RT "the mitochondrial genome of the mosquito Anopheles gambiae: DNA
 sequence, genome organization, and comparisons with mitochondrial
 sequences of other insects.";
 RL Insect Mol. Biol. 2:103-104(1993).
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
 DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L20934; AAD12194.1;
 DR InterPro: IPR000568; ATP_synA.
 DR Pfam: PF00119; ATP-syntA.1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASE_A.1.
 KW Hydrogen Ion transport; CF(0); Mitochondrion; Transmembrane.
 SQ SEQUENCE 226 AA; 25276 MW; 0AE0AB88A78AABAD CRC64;

Query Match 5.8%; Score 96.5; DB 1; Length 226;
 Best Local Similarity 23.4%; Pred. No. 0.63;
 Matches 59; Conservative 38; Mismatches 98; Indels 57; Gaps 11;

QY 49 LISNLFVPEFLOSLSSTSHISFK--DAFLGILIPGYAALPNPQ--FSQFVKDLTK 104
 Db 1 MMTNLFSV---FPPSTIILSNLSTFGLFLIVSVYMLNRRQVIMNNLLTLHK 56
 QY 105 CVAMLL--TLXGCVLDV-LYHLNPKSSILEDPEHEFLNIWSEFRNFAPITEEIFYT 161
 Db 57 EFKTLGSPGSHNGSTLMEISLFSLI-----MENNFLGLFPY-----1FTS 96
 QY 162 SMLTLYNLNLPHSOLSQOLFQPSLFFGLAHNAHAYEQLQBSMTTVSILLTTCFOIL 221
 Db 97 TSHLTFLALFPLMLSLMYG-----INRQHEFAHLVPGCTPPVLPMPVNCJETI 149
 QY 222 YTLFGGLTKEFVFRGTGNLMCCIIILHALCNMGFPGPSRLNHFVYVKKAGRISKVLS 281
 Db 150 SNVIRPGF---LAVRLTANM---TAGHLLTLTLGNTGPMASNY-----LILS 190
 QY 282 IWNKCYFALLV 293
 Db 191 LILTTQIALLV 202

RESULT 9

ID Y255_MYCPN STANDARD; PRT; 534 AA.
 AC P75422;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN MG255 HOMOLOG (H91-ORF534).
 GN MPN358 OR MP478.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
 RA Hermann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 pneumoniae.";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000047; AAB96126.1;
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 SQ SEQUENCE 534 AA; 62643 MW; 08B95C4D515FA6BC CRC64;

Query Match 5.8%; Score 96.5; DB 1; Length 534;
 Best Local Similarity 26.3%; Pred. No. 1.6;
 Matches 56; Conservative 35; Mismatches 85; Indels 37; Gaps 11;
 QY 3 QFSTFVLVLYISIVLPVATSOPEGSKRDNPTIKSRMOKTILISNLFVPLQSQ 62
 Db 270 QAMNLLLLKQTFNCFV-----VCCKESNMILLINKVPEPDINKOSSLYL--DKSQ 319
 QY 63 LSSITSHISFKDAFLGII-----IPGYAALPNPQFSQF---VVDLTKCVAMLLTLY 113
 Db 320 ISPLAQISKYNLLFEELALADMFYLEDFFALLKTPQIVNLFERIKQNLEKFPQPTLV 379
 QY 114 CGPVLDFV-LYHLNPKSSILEDPEHEFLNI-WSEFRNFAPITEEIFYTSMLTTYLN- 170

Db 360 ----FDYALMALV-----IATDF--EFNNVLMSPNDYLSLNKKQEDYASVNLAFENR 428

QY 171 -LIPHSQSLSTOOLWQPSLFPGLAHAAHAEOL 202

Db 429 SLKNHKDMS---LLEKPELFTMHAYIPETTYOV 458

RESULT 10

GPRI_HUMAN STANDARD: PRT: 331 AA.

AC Q14330;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROBABLE G PROTEIN-COUPLED RECEPTOR GPRI8 (FRAGMENT).

GN GPRI8 OR GPCR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-9734914; PubMed-9205118;

RA Gantz I., Murack A., Yang Y.-K., Samuelson L.C., Zimmerman E.M.,

RA Cook H., Yamada T.;

RT "Cloning and chromosomal localization of a gene (GPRI8) encoding a

RT novel seven transmembrane receptor highly expressed in spleen and

RT testis.";

RL Genomics 42:462-466(1997).

CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN TESTIS AND SPLEEN.

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: L42324; AAB65819.1; -

DR GCRDB: GCR_1864; -

DR MIN: 602042; -

DR InterPro: IPR000276; GPCR_Rhodopsn.

DR Pfam: PF00001; 7tm_1; 1.

DR PROSITE: PS00237; G_PROTEIN_RECPT_FL_1; 1.

DR PROSITE: PS50262; G_PROTEIN_RECPT_FL_2; 1.

KW G-protein coupled receptor; Transmembrane; Glycoprotein.

FT DOMAIN 1 26

FT TRANSSEM 27 47

FT TRANSSEM 48 56

FT TRANSSEM 57 77

FT TRANSSEM 78 95

FT TRANSSEM 96 116

FT TRANSSEM 117 138

FT TRANSSEM 139 159

FT TRANSSEM 160 191

FT TRANSSEM 192 212

FT TRANSSEM 213 232

FT TRANSSEM 233 253

FT TRANSSEM 254 268

FT TRANSSEM 269 289

FT TRANSSEM 290 331

FT CARBOHYD 14 14

FT NON_TER 331 331

SO SEQUENCE 331 AA; 38061 MW; 95C9C33EE25D7606 CRC64;

Query Match 5.8%; Score 96; DB 1; Length 331;

Best Local Similarity 24.3%; Pred. No. 1;

Matches 37; Conservative 31; Mismatches 36; Indels 48; Gaps 9;

QY 7 FLVLLYIS-S-YVPLVTSQDEGSKRNPRTIKSRMOKLTIMLSNFL-----VPL 59

Db 196 FLIPFLFMIGCYLVIIHMLHGRTSK-----LKEPKVEKRSIRIITLLVGLVCFMPF- 248

QY 60 QSLSTSTSHISFKDAFLGIIIPGYAALPNP-WQSFQVVDLTKCYAMLLTLGCPVL 118

Db 249 -----HICE-AFLMLGTGENSY---NPKAFITFLMLNSTC-----L 281

QY 119 DEVLHMLPKRSILEDYHEFLNIMSPNFI 150

Db 282 DVLILYIYS-----KQFQARIYSVMLRYNL 307

RESULT 11

YA35--SCHPO STANDARD: PRT: 426 AA.

ID YA35_SCHPO

AC Q09712;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHETICAL 48.8 KDA PROTEIN C18B11.05 IN CHROMOSOME 1.

GN SPAC18B11.05.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_Taxid=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: TO YEAST YBR004C AND SOME, TO C.ELEGANS ZK1321.3.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: Z50728; CAA90590.1; -

DR KW Hypothetical protein; Transmembrane.

FT TRANSSEM 33 53

FT TRANSSEM 104 124

FT TRANSSEM 129 149

FT TRANSSEM 182 202

FT TRANSSEM 223 243

FT TRANSSEM 264 284

FT TRANSSEM 319 339

FT TRANSSEM 351 371

FT TRANSSEM 403 423

SO SEQUENCE 426 AA; 48824 MW; CAC657F2ADA80FC CRC64;

Query Match 5.8%; Score 96; DB 1; Length 426;

Best Local Similarity 21.9%; Pred. No. 1.4; Mismatches 93; Indels 74; Gaps 12;

Matches 58; Conservative 40; Mismatches 93; Indels 74; Gaps 12;

QY 11 LYSISVPLVATYTSQPSGSKRDNPTIKSRMOKLTIMLSNFLVPLQSLSTSTSHI 70

Db 179 IYMSVGYTESLFAARSPFG-----LLLFIKQQYPAALFMSLATILRSNG 223

QY 71 SFKDAFLGIGITPGYALPNP-----WQSFQVVDLTKCYAMLLTLGCPVLDFVLY--- 123

Db 224 IFWCIFFGMPAIGTLKISLERLQLTFTFMQVSQLVGTGTCILTVFFFYNOYLGKFLCPG 283

QY 124 -----HLNPKRSILEDYFH-BFLNIMWSR---NFIFFA-----PITEEFY--TSM 164

Db 284 VAMCKSLPLIYP--AVOEKYMVGFRLRYMTLNINPFLALLSIIPILFALFYSISGST 341
 QY 165 LTTYLNLPHSOLSYQOLFQWOPSLFEGFLAHAAHEQLOEGSMYVSLTLTTCQILTT 224
 Db 342 LHSFRSISHLVLS--ALY---LYIGCFHMH-----TOVLNRMSSALPLLYWS 384
 QY 225 L-----FGLTKFEV 235
 Db 385 MAHATLVAKSRNLKAFGHCILFVMT 409

RESULT 12
 ATP6_DROSI STANDARD: PRT: 224 AA.
 ID ATP6_DROSI

AC P50269:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
 GN ATP6 OR ATPASE6.
 OS Drosophila simulans (Fruit fly).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7240;
 RX MEDLINE=93374296; PubMed=8365657;
 RA Kaneko M., Satta Y., Matsura E.T., Chigusa S.I.;
 RT "Evolution of the mitochondrial ATPase 6 gene in Drosophila:
 RT unusually high level of polymorphism in D. melanogaster.";
 RL Genet. Res. 61:195-204(1993).

CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: S649377; AAD13958.1;
 DR FlyBase: FBgn0012875; DslnMtc:ATPase6.
 DR InterPro: IPR000568; ATP_synth_A.
 DR Pfam: PF00119; ATP-synth_A.1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASEA.
 KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
 SQ SEQUENCE 224 AA; 25065 MW; 9480FE8215156782 CRC64;

Query Match 5.88; Score 95.5; DB 1; Length 224;
 Best Local Similarity 24.08; Pred. No. 0.75;
 Matches 61; Conservative 40; Mismatches 90; Indels 63; Gaps 15;

QY 49 LISNLVLPFLQSLSTSHSKDAFLGIGITPGYALPN---PWOSQEVKDLTK 104
 Db 1 MMTNLFVS-FDPSAIFNLS--LNMWSTFLGLMIPSIYMWPSKRYNIWN----- 47
 QY 105 CVAALLTYCGPVLDFVYHLNPKSSILDFYHEFLNWS---FRNFIPATEEIPYT 161
 Db 48 --SILLTLH---KEFTK--LLGPSGNHGSGTF--FISLEFLILFNFM--GLPIYITTS 95
 QY 162 SMLTTYLNL-IPHSOLSYQOLFQWOPSLFEG-LSAHAAHEQLOEGSMYVSLTLTTCQ 219

Db 96 TSHLTLTSLALP-----LMICFMYLGWINHTQGMFALHYVQGNPALIMPRMVIC 146
 QY 220 ILTYTLFEGFLKFEVFRGCMIMCCIIHALCNMGEPGSRNLHFTVVDKAKRISK 279
 Db 147 TISNIIRPT---LAVRLTANM---IAGHLTLTLTGNGPSMSYL-----L 186
 QY 280 VSIMNKCYPALLVL 293
 Db 187 VFLTLTAQIALLV 200

RESULT 13
 ATP6_DROYA STANDARD: PRT: 224 AA.
 ID ATP6_DROYA

AC P00851:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
 GN ATP6 OR ATPASE6.
 OS Drosophila yakuba (Fruit fly).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2317.6 IVORY COAST;
 RA MEDLINE=86089137; PubMed=3001325;
 RX Clardy D.O., Wolstenholme D.R.;
 RT "The mitochondrial DNA molecular of Drosophila yakuba: nucleotide
 RT sequence, gene organization, and genetic code.";
 RL J. Mol. Evol. 22:252-271(1985).

CC [2]
 CC SEQUENCE FROM N.A.
 CC STRAIN=2317.6 IVORY COAST;
 CC MEDLINE=83246544; PubMed=6306579;
 CC Clardy D.O., Wolstenholme D.R.;
 CC "Nucleotide sequence of a segment of Drosophila mitochondrial DNA
 CC that contains the genes for cytochrome c oxidase subunits II and III
 CC and ATPase subunit 6.";
 CC Nucleic Acids Res. 11:4211-4227(1983).

CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X03240; CAA26989.1;
 DR EMBL: X00924; CAA25442.1;
 DR PIR: G93477; PMFE6Y.
 DR PIR: E25797; E25797.
 DR FlyBase: FBgn0013177; DyakMtc:ATPase6.
 DR InterPro: IPR000568; ATP_synth_A.
 DR Pfam: PF00119; ATP-synth_A.1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASEA.
 KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
 SQ SEQUENCE 224 AA; 25141 MW; 27D6DBF5791FD24 CRC64;

FT	TRANSMEM	101	120	3	(POTENTIAL).
FT	DOMAIN	121	139	4 <td>CYTOPLASMIC (POTENTIAL).</td>	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	140	164	4 <td>(POTENTIAL).</td>	(POTENTIAL).
FT	DOMAIN	165	205	5 <td>EXTRACELLULAR (POTENTIAL).</td>	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	206	226	5 <td>(POTENTIAL).</td>	(POTENTIAL).
FT	DOMAIN	227	239	6 <td>CYTOPLASMIC (POTENTIAL).</td>	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	240	260	6 <td>(POTENTIAL).</td>	(POTENTIAL).
FT	DOMAIN	261	271	7 <td>EXTRACELLULAR (POTENTIAL).</td>	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	272	292	7 <td>(POTENTIAL).</td>	(POTENTIAL).
FT	DOMAIN	293	312	7 <td>CYTOPLASMIC (POTENTIAL).</td>	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	5	5	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	DISULFID	97	179	BY SIMILARITY.	
SO	SEQUENCE	312 AA;	35179 MW;	56D164BF9D59EA9	CRC64.
Query Match					
Best Local Similarity		5.7%;	Score 94.5;	DB 1;	Length 312;
Matches		60;	Conservative	55;	Mismatches 97;
					Indels 111;
					Gaps 16;
QY	46	TIMLSINFLVPEFLOSQSSSTT-----SHISKFAEFLGLGITIPGYAALPMPQSFQV	99		
DB	36	TITLTNTGLIALISIDLOQPMYIFLONSFTDA-----VSVIVIPKMLATFL	86		
QY	100	KDLKRCVAMLLTYCGPVLDFVLYHLLMPKSSILDEYHEPFINMSFRNFPAPTEEIF	159		
DB	87	EE-TKTTIS-----YVCGILOFYSEFVLLVRECLL-----LAVAYDR--YAAICKPPL	131		
QY	160	YTSML-----TYVNLNIPHS---QLSY-----QDLFMQPSLEFGIAHAH	196		
DB	132	YPAIMTKAVCMRWKLYKGLSIAPLNFNLVHTSGLKLSFSCSNVNVNPFQCNPSLPQISSSS	191		
QY	197	HAYEQLQE--GSMTTSVILLT-----TC-FQIIYTLF	226		
DB	192	TALNLELTFEIGSLPEVMSIIITLISYFILLTVVIRISKERKRYAFSTCTSHLMAVSLE	251		
QY	227	GLGTRFVAVRRGGINMCCILHALCNIMGFPSPSLNLHFTYVDKKAQISLVSIGW-NK	285		
DB	252	HGTVEMTFEQPANN-----PSLDKDKTIMSLFTYV---IPMLNPLIYSWRNK	295		
QY	286	CYFALLVGLISLKDTLQTLVGT	308		
DB	296	-----EVKDALHRAIAT	307		
RESULT 15					
ATP6_ANOOU STANDARD; PRt: 226 AA.					
AC	ATP6_ANOOU	STANDARD;	PRt:	226	AA.
DT	01-FEB-1994	(Rel. 28, Created)			
DT	01-FEB-1994	(Rel. 28, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).				
GN	ATP6.				
OS	Anopheles quadrimaculatus (Mosquito).				
OG	Mitochondrion.				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;				
CC	Anopheles.				
OX	NCBI_TaxID=7166;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ORLANDO;				
RX	MEDLINE=92190510; PubMed=2134168;				
RA	Cockburn A.F., Mitchell S.E., Seayright J.A.;				
RT	"Cloning of the mitochondrial genome of Anopheles quadrimaculatus.";				
RL	Arch. Insect Biochem. Physiol. 14:31-36(1992).				
CC	-I- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A				
CC	DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.				
CC	-I- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC				
CC	CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE				
CC	SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)				
CC	HAS THREE MAIN SUBUNITS: A, B AND C.				
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				

CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04272; AAA9354.1; -
CC InterPro: IPR000568; ATP_synth_A.
CC Pfam: PF00119; ATP-synth_A: 1.
CC PRINTS: PR00123; ATPASEA.
CC PROSITE: PS00449; ATPASE.A: 1.
CC Hydrogen Ion transport; CF(0); Mitochondrion; Transmembrane.
KW
SQ SEQUENCE 226 AA; 25286 MW; FA4989EDAEFF1571 CRC64;

Query Match 5.7%; Score 94; DB 1; Length 226;
Best Local Similarity 23.9%; Pred. NO. 0.99; Mismatches 44; Gaps 10;
Matches 53; Conservative 35; Indels 90; Indels 44; Gaps 10;
QY 49 LISNLFVLPFLOQSSTSHISFK--DAFLGILIPGYAALPNWQ--FSQFVKDLTK 104
DB 1 MMTNLFSEV---FDPSTTILNLSNWLSTFLGILLIPSEFWLPPNRQVWNNILLTLHK 56
QY 105 CVAMLL--TYCCPVLDV--LYHLNPKSSILEDYHEFLNIMSFNFIAPITEETPYT 161
DB 57 EFKTLGSPSGHNGSTLMFISLFSLI-----MFNNFLGLFPY-----IFTS 96
QY 162 SMLTTYLNIPHSQLYQQLFQMOPSLFFGLAHNAHAYEOLQESMTTVSILLTFCQIL 221
DB 97 TSHLTTLTALAFLPLWLSFMLYGW-----INHTQHMAHLVPOGTPPVLMPPWVCJETI 149
QY 222 YTLFGGLTKRVFVRTGNNLCCITLHALCNIMGPPSPRLN 263
DB 150 SNVIRPGT---LAVRLTANM---IAGHLLLTLLGNTGPMPTTN 185

Search completed: February 12, 2002, 09:32:25
Job time: 341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 11:09:22 : Search time 277.71 Seconds
(without alignments)
5634.002 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825
Sequence: 1 accctacccctttttctatct.....aaataacagattaacatt 1825

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq_1101:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1802	98.7	1825	19	AAV17601	Yeast Afc1 gene en
2	140.4	7.7	1616	21	AA36991	Arabidopsis thalia
3	137	7.5	936	22	AA58252	Oligonucleotide D1
4	137	7.5	936	22	AA58254	Oligonucleotide D1
5	137	7.5	936	22	AA58257	Oligonucleotide D1
6	137	7.5	936	22	AA58259	Oligonucleotide D2
7	137	7.5	936	22	AA58262	Oligonucleotide D2
8	137	7.5	936	22	AA58255	Oligonucleotide D1
9	136.8	7.5	936	22	AA58252	Oligonucleotide D1
10	136.8	7.5	936	22	AA58254	Oligonucleotide D1
11	136.8	7.5	936	22	AA58257	Oligonucleotide D1

12	136.8	7.5	936	22	AA58259	Oligonucleotide D2
13	136.8	7.5	936	22	AA58262	Oligonucleotide D2
14	136.8	7.5	936	22	AA58255	Oligonucleotide D1
15	117.4	6.4	2966	20	AA83493	Human farneylated
16	117.4	6.4	2966	20	AAV83752	CDNA encoding huma
17	106	5.8	2570	21	AA58173	Lung cancer associ
18	88.2	4.8	684	21	AA58173	Trichoderma reesei
19	79.8	4.4	649	21	AA58173	Fusarium venenatum
20	72.2	4.0	844	22	AA58173	Human CDNA clone (
21	72.2	4.0	2161	22	AA58173	Human CDNA sequenc
22	66.8	3.7	1072	22	AA58173	H. pylori HPC187 e
23	63.6	3.5	1245	22	AA58173	H. pylori HPS187 e
24	61.6	3.4	1224	18	AA58173	H. pylori cell env
25	56.6	3.1	244	22	AA58173	Oligonucleotide D1
26	56.2	3.1	310	22	AA58173	Oligonucleotide D1
27	53.8	2.9	464	22	AA58173	Human cervical can
28	53	2.9	7204	20	AA58173	Human colon cancer
29	52.8	2.9	1673	20	AA58173	O. longistaminta X
30	52.2	2.9	244	22	AA58173	Human secreted pro
31	51.8	2.8	12020	21	AA58173	Oligonucleotide D1
32	51.6	2.8	1641	17	AA58173	ARSDRI polynucleot
33	51.4	2.8	382	15	AA58173	Human interleukin-
34	51.4	2.8	905	21	AA58173	Grapevine ribosoma
35	50.8	2.8	612	22	AA58173	Partial sequence M
36	50	2.7	797	22	AA58173	Human cervical can
37	49.8	2.7	788	22	AA58173	Human secreted pro
38	49.6	2.7	1024	21	AA58173	Human colon cancer
39	49.6	2.7	1459	22	AA58173	Human prostate can
40	49.2	2.7	568	22	AA58173	Maize RAD51 orthol
41	49.2	2.7	568	22	AA58173	Probe #9087 for ge
42	49.2	2.7	621	22	AA58173	Human cervical can
43	49.2	2.7	2040	22	AA58173	Human polynucleoti
44	49	2.7	887	21	AA58173	Human secreted pro
45	48.8	2.7	208	21	AA58173	Human colon cancer

ALIGNMENTS

RESULT 1	AAV17601	standard; DNA: 1825 BP.
ID	AAV17601	
XX	AAV17601:	
AC	20-JUL-1998	(first entry)
XX	Yeast Afc1 gene encoding Afc1 protein.	
DE	Yeast Afc1 gene encoding Afc1 protein.	
XX	AFc1 gene; Afc1p, a-factor convertase; CNA protease;	
KW	zinc metalloprotease; yeast; prenylation; Ras; inhibitor; cancer;	
KW	colorectal carcinoma; pancreas carcinoma; leukemia; therapy; ss.	
XX	Saccharomyces cerevisiae.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	CDS	342..1704
FT		/*lag= a
FT		/transl_except= (pos:378..380, aa:Thr)
FT		/transl_except= (pos:702..705, aa:Ser)
FT		/note= "this codon has an apparent 1 nucleotide
FT		insertion, which alters the reading
FT		frame"
FT		/transl_except= (pos:1555..1557, aa:Met)
FT		/transl_except= (pos:1663..1665, aa:Leu)
XX	W09805786-A2.	
XX	12-FEB-1998.	
XX	06-AUG-1997:	97WO-US14777.
XX	30-JUL-1997:	97US-0902774.

PR 07-AUG-1996; 96US-0023491.
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Ashby MN, Boyarchuk VL, Rine JD;
 PI
 XX WPI: 1998-145621/13.
 DR P-PSDB: AAM48301.
 XX
 PT Vector comprising nucleic acid coding for Afcip and Rceip proteins
 PT involved in protein prenylation - useful for screening for
 PT inhibitors of proteins, especially of mutated ras proteins involved
 PT in cancerous conditions
 XX
 PS
 XX
 CC Claim 11: Page 47-48; 59pp: English.
 CC
 CC This DNA sequence comprises the yeast Afc1 (a-factor convertase)
 CC gene that codes for a novel protein (see AAM48301), designated Afc1p,
 CC that mediates the removal of AAX tripeptide from CAAX proteins
 CC following prenylation. Afc1 was isolated using an autoactive
 CC arrest, sensitised selection for CAAX proteases. This involved
 CC ecotopic expression of an a-factor in an a-cell and a CAAX
 CC permutational defective for proteolysis. Of 127 mutants isolated, 22
 CC had mutations in a single gene (AFC1). This is the first CAAX
 CC protease to be identified. Genetic knockout of Afc1 resulted in
 CC incomplete reduction of a-factor production, thereby exposing the
 CC presence of a second CAAX protease. The gene for this (see AA17602)
 CC was named RCE1 (Ras and a-factor converting enzyme). Vectors
 CC comprising the Afc1 or Rce1 gene, polypeptides encoded by such
 CC vectors and recombinant cells transduced with the vectors are
 CC claimed. The novel Afc1p and Rce1p (see AAM48302) proteins can be
 CC used to screen for inhibitors of prenylation which can be used
 CC in the treatment of Ras dependent cancers such as colorectal and
 CC exocrine pancreatic carcinomas or myeloid leukaemia.
 CC
 XX
 SQ Sequence 1825 BP: 570 A; 398 C; 280 G; 577 T; 0 other:
 Query Match 98.7%; Score 1802; DB 19; Length 1825;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1824; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 481 ccactgtgctggaagacgaatgatgatgaacttttcaataatcaaggacactcc 540
 |||||||
 Db 480 ccaactgtgctggaagacgaatgatgatgaacttttcaataatcaaggacactcc 539
 QY 541 cgggccaaggccaatttccatttcgtgagcgtctaaactgaagccaagctagt 600
 |||||||
 Db 540 cgggccaaggccaatttccatttcgtgagcgtctaaactgaagccaagctagt 599
 QY 601 ttcatcaataagacactttcccttaaatctggcagatgcccgtttcttattgaa 660
 |||||||
 Db 600 ttcatcaataagacactttcccttaaatctggcagatgcccgtttcttattgaa 659
 QY 661 gtcccgacgacgaattcatatgtgtccactgtgcgacagag-ttaigtcttctgg 719
 |||||||
 Db 660 gtcccgacgacgacattcatatgtgtccactgtgcgacagagtttaigtcttctgg 719
 QY 720 tctctatccagttgtgtactctgtgtgatttggcaactctctactaataagcattt 779
 |||||||
 Db 720 tctctatccagttgtgtactctgtgtgatttggcaactctctactaataagcattt 779
 QY 780 cctggaagaaaatttggtttccaataattgacggtccaactaagatccagatagat 839
 |||||||
 Db 780 cctggaagaaaatttggtttccaataattgacggtccaactaagatccagatagat 839
 QY 840 caagagctgactgtggtatgctatgtgtggcccaactcttaccgttcccttaagat 899
 |||||||
 Db 840 caagagctgactgtggtatgctatgtgtggcccaactcttaccgttcccttaagat 899
 QY 900 ctgtgataaattccctactgatttcccttggtaactaagttctctgttctgtgtcca 959
 |||||||
 Db 900 ctgtgataaattccctactgatttcccttggtaactaagttctctgttctgtgtcca 959
 QY 960 aacttgaacatgacacattccagcttccatcatatgcccagtgttaataagttcaccc 1019
 |||||||
 Db 960 aacttgaacatgacacattccagcttccatcatatgcccagtgttaataagttcaccc 1019
 QY 1020 atggaagacggtgacacgaataaattcattgaaagtgtggccgaagagttgtccc 1079
 |||||||
 Db 1020 atggaagacggtgacacgaataaattcattgaaagtgtggccgaagagttgtccc 1079
 QY 1080 tctagataagaatttgttcatctgacggtcctaaagaatcttccatccaacgcatatt 1139
 |||||||
 Db 1080 tctagataagaatttgttcatctgacggtcctaaagaatcttccatccaacgcatatt 1139
 QY 1140 cagaggttgcacattccactccaagagaattgttgttgcacacttaagtgacaagtaa 1199
 |||||||
 Db 1140 cagaggttgcacattccactccaagagaattgttgttgcacacttaagtgacaagtaa 1199
 QY 1200 ttctactgatgaattacggtgtgttggccatgaaatcggtcactggaataaacc 1259
 |||||||
 Db 1200 ttctactgatgaattacggtgtgttggccatgaaatcggtcactggaataaacc 1259
 QY 1260 catcgttaataatgttcatctttagtcaatttgcacacttccatcttccctttcac 1319
 |||||||
 Db 1260 catcgttaataatgttcatctttagtcaatttgcacacttccatcttccctttcac 1319
 QY 1320 cagcatctacagaataatcatatttacaacacttggcgttttctttagagaagttcac 1379
 |||||||
 Db 1320 cagcatctacagaataatcatatttacaacacttggcgttttctttagagaagttcac 1379
 QY 1380 tggcagtttgttgaatccggttatcatcaagaagaattccccaattcatgtgattatgt 1439
 |||||||
 Db 1380 tggcagtttgttgaatccggttatcatcaagaagaattccccaattcatgtgattatgt 1439
 QY 1440 atttaagactatttaactccacactcgaatgagcagatgcaatctgtgattgaattatctc 1499
 |||||||
 Db 1440 atttaagactatttaactccacactcgaatgagcagatgcaatctgtgattgaattatctc 1499
 QY 1500 cagaactcatgaatatcaagctgattgtcttaaaaaatttgggtctacaagcaaaatct 1559
 |||||||
 Db 1500 cagaactcatgaatatcaagctgattgtcttaaaaaatttgggtctacaagcaaaatct 1559

Oy 1560 atgtaggctctaatgtactacaatacaaaaccttccaccatgtagatcctct 1619
|||||
Db 1560 atgtaggctctaatgtactacaatacaaaaccttccaccatgtagatcctct 1619
Oy 1620 gtatctagctatcatatttcccatccaacttagctgaagaatgcgcgtctgagacta 1679
|||||
Db 1620 gtatctagctatcatatttcccatccaacttagctgaagaatgcgcgtctgagacta 1679
Oy 1680 tgttgtaaaagaagaanaactatctatagatgacatatttagcgtgtacgcttaaa 1739
|||||
Db 1680 tgttgtaaaagaagaanaactatctatagatgacatatttagcgtgtacgcttaaa 1739
Oy 1740 ttccagcttcgtatgtctacatactacacacacaggtatctactataaagaatgaag 1799
|||||
Db 1740 ttccagcttcgtatgtctacatactacacacacaggtatctactataaagaatgaag 1799
Oy 1800 aaagaanaataacgattaaacatt 1825
|||||
Db 1800 aaagaanaataacgattaaacatt 1825

RESULT 2
AAC36991
ID AAC36991 standard; DNA; 1616 BP.
XX AAC36991;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 15792.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 23-JUL-1999; 99US-0145222.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.


```

XX 24-APR-2001 (first entry)
DT .....
XX Oligonucleotide D2007.
DE .....
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS .....
XX WO200107665-A2.
PN .....
XX 01-FEB-2001.
PD .....
XX 26-JUL-2000; 2000WO-US20476.
PF .....
XX 26-JUL-1999; 99US-0145695.
PR .....
XX 17-MAR-2000; 2000US-0190259.
XX .....
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA .....
XX Umek RM;
PI .....
XX WPI: 2001-159728/16.
DR .....
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX .....
XX Example 6; Page 128; 159pp; English.
PS .....
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX .....
XX Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
SQ .....

```

```

Query Match 7.5%; Score 137; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 6.6e-23;
Matches 11; Conservative 499; Mismatches 289; Indels 0; Gaps 0;

OY 1023 gagagacgtgacacgaaataatcatgaaagttggccgataagttggtccctc 1082
DB 800 GCGGGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 741
OY 1083 agataagattttgtcatgacgctcaaaaagatctctcattcaaacgataattcac 1142
DB 740 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 681
OY 1143 aggttgcattaccccaagagaatgtttgttcgacattagtagaagaatttc 1202
DB 680 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 621
OY 1203 tactgatgaattacgctgttttggcccatgaaatcgctcactgcaaaaaaacacat 1262
DB 620 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 561
OY 1263 cgtataatgcatctttagtcaattgacacctctcctcttctcccttaccag 1322
DB 560 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 501
OY 1323 catctacagaatacatcatttacaacaccttcgcttcttagagaagtcactgg 1382
DB 500 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 441
OY 1383 cagtttctgaccccgatcaccaagaatcccatcatcatgattatgtattc 1442
DB 440 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 381

```

```

OY 1443 taagcatttaactccactcgcgaatggtccatgcaattcgtagttagttaccag 1502
DB 380 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 321
OY 1503 aactcagaaatacgaagctgactatgcttaaaatgggctacaaagcaaatctag 1562
DB 320 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 261
OY 1563 tagggcttaattgctcacaataacaaaccttcacacatgaaatgtagtcctcgt 1622
DB 260 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 201
OY 1623 ttctagctatcatcttccactccactcgaagatcgaccgctctagactatgt 1682
DB 200 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 141
OY 1683 tagtgaagaagaagaactaactatagagtagacatatagcattgacgttaattc 1742
DB 140 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 81
OY 1743 agcttcgttatgtcatatcatcacatcacacaggtatctactataagaataagaa 1802
DB 80 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 21
OY 1803 gaaaaataaacgattaaa 1821
DB 20 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 2

```

```

RESULT 8
AAF58255/C
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
DB 01-FEB-2001.
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX WPI: 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX .....
XX Example 6; Page 127; 159pp; English.
PS .....
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX .....
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SQ .....

```


DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX

QY	1034	aacggaataatctatgtgaagtgtgcccagatagatgtggtctccctctagataagattc	1093
Db	8	www.....	67
QY	1094	tgtcatctgcagctcaaaaagatctcttcattccaacgcatattccagatttgcatt	1153
Db	68	www.....	127
QY	1154	tcaccccaagaaatgttttctgcacactttagtgaacgtaattctactgtgaa	1213
Db	128	www.....	187
QY	1214	ttacgctgttcttgcacatgaatctgcactgcgcaaaaacacatcgttaattcgy	1273
Db	188	www.....	247
QY	1274	tcactttagtaattgcacaccttcattctccctttcacacgactacagaa	1333
Db	248	www.....	307
QY	1334	atcacatatttacaaccttcgcttctctttagagaagtcacactgcagtttctgy	1393
Db	308	www.....	367
QY	1394	atcccggtatcactaagaatcccatatataatgattatgttatttaagactat	1453
Db	368	www.....	427
QY	1454	taactcaactcgaattgcacatgaacttgatgatttaattccagaactcagat	1513
Db	428	www.....	487
QY	1514	atcaagctgatctatctgcttaaaaaatgttgctacagaacaaatctatgttggctctaa	1573

```
Db 488 .....
Oy 1574 ttgattcacaatacaaaacttccaccatgagtgtgacccctgtatctgactc 1633
Db 548 .....
Oy 1634 attatcccatccactctagctgaagaatcgacccgtctgactgtgtgaaaga 1693
Db 608 .....
Oy 1694 agaaaactaactatagatgagacatatagatgacgttaaatcagcttgat 1753
Db 668 .....
Oy 1754 gtctatatctacacatacaggtatctactataagaataaagaaataaa 1813
Db 728 ..... 787
```

RESULT 10

AAF58254
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry).

DE Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

XX WO200107665-A2.

PD 01-FEB-2001.

PE 26-JUL-2000; 2000MO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match

Best Local Similarity 0.8%; Score 136.8; DB 22; Length 936;

Matches 6; Conservative 496; Mismatches 278; Indels 0; Gaps 0;

```
Oy 1034 aactgaataatctatgtaaggttgccgtagagtggttccctagataaagatt 1093
Db 8 ..... 67
```

```
Oy 1094 ttgcatgtacgctcaaaagatctctcatlccaaagcatattccaggttgcat 1153
Db 68 ..... 127
Oy 1154 tcacctcaagaagatttctgttcgacacttagtgaacagtaattctactgaa 1213
Db 128 ..... 187
Oy 1214 ttacggtgttcttgccatgaatcgltacgtgcaaaaaaacacatcgttaatg 1273
Db 188 ..... 247
Oy 1274 tcatcttagtcaattgcacaccttcctcattcttccttccacttcaagcactcaagaa 1333
Db 248 ..... 307
Oy 1334 atacatcatltaaacaccttcgcttcttcttagaagaatccactgacgttctg 1393
Db 308 ..... 367
Oy 1394 atccgtatacctaagaattcccatlcatlgtatgtatlttaacgactat 1453
Db 368 ..... 427
Oy 1454 taactcactcgaatggtccatgcaattcgatgagtttaattccaagaactgaat 1513
Db 428 ..... 487
Oy 1514 atcaagctgattctatgctaaatbgtgctacaagcaaatcattgtaggctcaa 1573
Db 488 ..... 547
Oy 1574 ttgattcaaaaaccaaaacttccacatgaattgagatcctctgatatctagctac 1633
Db 548 ..... 607
Oy 1634 attatcccatccactctagctgaagaatcgacccgtctgactagttagtgaaga 1693
Db 608 ..... 667
Oy 1694 agaaaactaactatagatgagacatatagatgacgttaaatcagcttgat 1753
Db 668 ..... 727
Oy 1754 gtctatatctacacatacaggtatctactataagaataaagaaataaa 1813
Db 728 ..... 787
```

RESULT 11

AAF58257
ID AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

XX WO200107665-A2.

PD 01-FEB-2001.

PE 26-JUL-2000; 2000MO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.


```
Oy 1274 tcaatttagtcaattgacacacctctcatttctcccttccacagcactacagaa 1333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 www. .... 307
Oy 1334 ataccattttacacaccttcgcttttcttagaagaatccctgcgcttctgtg 1393
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 www. .... 367
Oy 1394 atccgttatacctaagaattcccatatcatgtgattatgtatlaagcactat 1453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 www. .... 427
Oy 1454 taactccactgcgaatgctgcacatgcattgtagttagtlaattccagaactcgaat 1513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 www. .... 487
Oy 1514 atcaagctgcgactatgcttaaaaaatcggctacagaacaaatctatgtaggctctaa 1573
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 www. .... 547
Oy 1574 ttgactcaaaatcaaaaacctttccacatgtagatgcctctgtatctctatc 1633
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 www. .... 607
Oy 1634 attatcccatccaaactctagctgaagaatcgaccgctctagactatglttagaaga 1693
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 www. .... 667
Oy 1694 agaaaaactaatctataagtaacatatctagcatgtaacgttaaatcagcttcgttat 1753
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 668 www. .... 727
Oy 1754 gtctatatcatatacatacaaggtatctactataagaataagaaagaaaaataaa 1813
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 728 www. .... 787

RESULT 13
AAF58262
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
XX
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX
PS Example 6; Page 128; 159pp; English.
```

```
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
XX
Query Match 7.5%; Score 136.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 7.4e-23;
Matches 6; Conservative 496; Mismatches 278; Indels 0; Gaps 0;
Oy 1034 aacgaaaaaatctatgaaagtttggcgcgtagagttgggttccctatagaatatt 1093
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 www. .... 67
Oy 1094 ttgcatgacggtcacaagaatctctcatctcaaacgacatctcagaattgacct 1153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 www. .... 127
Oy 1154 tcacctcaagaagattgtttgttcgacacttagtgaacagtaattctactcgtatgaa 1213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 www. .... 187
Oy 1214 ttacgctgttttgcccatgaatcgctcgtgcgcaaaaaccacatcgtaataagg 1273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 www. .... 247
Oy 1274 tcaatttagtcaattgacacacctctccatcttctccctttccacagcactacagaa 1333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 www. .... 307
Oy 1334 ataccattttacacaccttcgcttttcttagaagaatccctgcgcttctgtg 1393
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 www. .... 367
Oy 1394 atccgttatacctaagaattcccatatcatgtgattatgtatlaagcactat 1453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 www. .... 427
Oy 1454 taactccactgcgaatgctgcacatgcattgtagttagtlaattccagaactcgaat 1513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 www. .... 487
Oy 1514 atcaagctgcgactatgcttaaaaaatcggctacagaacaaatctatgtaggctctaa 1573
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 www. .... 547
Oy 1574 ttgactcaaaatcaaaaacctttccacatgtagatgcctctgtatctctatc 1633
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 www. .... 607
Oy 1634 attatcccatccaaactctagctgaagaatcgaccgctctagactatglttagaaga 1693
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 www. .... 667
Oy 1694 agaaaaactaatctataagtaacatatctagcatgtaacgttaaatcagcttcgttat 1753
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 668 www. .... 727
Oy 1754 gtctatatcatatacatacaaggtatctactataagaataagaaagaaaaataaa 1813
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 728 www. .... 787

RESULT 14
AAF58255
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
```

DT	24-APR-2001	(first entry)
XX	Oligonucleotide D1876.	
DE		
XX	Electron-transfer group: ETM; mismatch; genotyping;	
RW	gene expression; ss.	
XX		
OS	Synthetic.	
XX		
PN	W0200107665-A2.	
XX		
PD	01-FEB-2001.	
XX		
PF	26-JUL-2000; 2000WO-US20476.	
XX		
PR	26-JUL-1999; 99US-0145695.	
XX		
PR	17-MAR-2000; 2000US-0190259.	
XX		
PA	(CLIN-) CLINICAL MICRO SENSORS INC.	
XX		
PI	Umek RM;	
XX		
DR	WPI; 2001-159728/16.	
XX		
PT	Nucleic acids containing electron-transfer group, useful as labels in	
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on	
PT	a single surface	
XX		
PS		
XX	Example 6; Page 127; 159pp; English.	
XX		
CC	The present invention relates to a composition comprising two nucleic	
CC	acids each containing an electron-transfer group (ETM) having	
CC	different redox potentials. The invention is used for electronic	
CC	detection of nucleic acids, especially of substitutions (mismatches)	
CC	and single-nucleotide polymorphisms, e.g. for genotyping,	
CC	monitoring gene expression.	
XX		
XX	Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;	
SO		

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:13:07 ; Search time 2973.29 Seconds

(without alignments)
6595.739 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825
Sequence: 1 acctaccctttttctatct.....aaataaacgattaacatt 1825

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: 9b_est1:*
11: 9b_est2:*
12: 9b_hic:*
13: 9b_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_hiv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	283.8	15.6	819	13	CNS06Y63
2	275.8	15.1	985	13	CNS07772
3	267	14.6	516	13	AZ925942
4	259.8	14.2	328	13	AZ923982
5	238.8	13.1	893	13	CNS06CIS
6	213.6	11.7	517	13	AZ930382
7	177.2	9.7	1116	13	CNS06CBM
8	161.6	8.9	1081	13	CNS06OC9
9	108.4	5.9	859	11	BGI76721
10	106.8	5.9	424	10	AA965341
11	105.6	5.8	543	11	BF051006
12	102.8	5.6	708	10	AM174887

13	97	5.3	394	11	T37813	T37813	EST103073	S
14	96.8	5.3	371	10	AA787399	AA787399	n3b02a1.r	
15	95.6	5.2	882	13	CNS01F15	AL141258	Anopheles	
16	95.4	5.2	586	10	AM146879	AM146879	614089B06	
17	92.4	5.1	651	11	BG557737	BG557737	EM1_56_C1	
18	92.2	5.1	491	10	BE459882	BE459882	EST415174	
19	92	5.0	493	11	BG556491	BG556491	EM1_37_G1	
20	90.6	5.0	589	10	AM585792	AM585792	EST317415	
21	87	4.8	430	10	AM759861	AM759861	8155a11.y	
22	84	4.6	531	10	BE054474	BE054474	GA_Ea001	
23	83.4	4.6	627	10	BE205165	BE205165	EST397841	
24	83.2	4.6	457	11	BG651426	BG651426	sad46e11.	
25	82.2	4.5	519	10	AI328626	AI328626	asd04ne.f	
26	81.2	4.4	388	10	AM759000	AM759000	sl33d11.y	
27	81.2	4.4	748	11	BG836625	BG836625	Zm08_0390	
28	80.6	4.4	472	10	AI330202	AI330202	c4d10ne.f	
29	80.6	4.4	648	10	AI066958	AI066958	HOT-10 C4	
30	78.6	4.3	648	11	BG874694	BG874694	RC2-BN007	
31	78	4.3	699	10	AI759788	AI759788	shd4c03.y	
32	75.2	4.1	471	10	AM678001	AM678001	WS1_12_H0	
33	75.2	4.1	499	10	AM677942	AM677942	WS1_12_H0	
34	75.2	4.1	871	13	AZ685824	AZ685824	ENTH54TF	
35	73.6	4.0	706	10	AI759796	AI759796	shd4d03.y	
36	73.4	4.0	600	10	AM520069	AM520069	660056F02	
37	73.2	4.0	699	11	BG862090	BG862090	602795627	
38	72.8	4.0	280	10	AM120333	AM120333	614089B05	
39	72.8	4.0	440	10	AI757461	AI757461	ELSTGa05	
40	72.4	4.0	1224	10	BE035468	BE035468	MO06D10 M	
41	71	3.9	446	10	AM459275	AM459275	sh22h03.y	
42	70.8	3.9	603	10	BE060801	BE060801	HVMEg001	
43	65.4	3.6	609	10	AU037739	AU037739	AU037739	
44	65.4	3.6	805	13	CNS009F8	CNS009F8	Drosophila	
45	65	3.6	365	10	AL515373	AL515373	AL515373	

ALIGNMENTS

RESULT 1	CNS06Y63	819 bp	DNA	GSS	06-JUL-2001
LOCUS	T7 end of clone AY00A006E12 of library AY00A	from strain CBS 6340			
DEFINITION	of Kluyveromyces thermotolerans, genomic survey sequence.				
ACCESSION	AL420625				
VERSION	AL420625.1	GI:12203813			
KEYWORDS	GSS.				
SOURCE	Kluyveromyces thermotolerans.				
ORGANISM	Kluyveromyces thermotolerans				
REFERENCE	1 (bases 1 to 819)				
AUTHORS	Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 10.				
JOURNAL	FEBS Lett. 487 (1), 61-65 (2000)				
MEDLINE	20584720				
REFERENCE	2 (bases 1 to 819)				
AUTHORS	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., de Montigny,J., Dujon,B., Durieux,P., Casaregola,S., Saurin,W., Tekala,F., Toffano-Nicche,C., Mesolowski-Louvel,M., Wincker,P. and Weissbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)				
MEDLINE	20584711				
REFERENCE	3 (bases 1 to 819)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :				

OY	1342	tttaaacacacttcggcttttcttaagaagtcaccatgcagttt-----tgct	1392
Db	185	CTTTACATTTATTTGGATTCTCTTTGGGTGGCCCACTGGAGATTACTCTCAAGTTCA	244
OY	1393	gaccgcgtatacctaagaagatccccaatcaatctgattatgltatlttaacagctta	1452
Db	245	AAGCAAGCTTTACACCGGAAHTGCCCTATCATGATGTGGCTTACATTTGTTAAATGATTTA	304
OY	1453	ttaactcaccctcgaattgycatgcaatctgtatgagtttaatttcagaactatgaa	1512
Db	305	TTGGCCCAATGGATGGGTGATGCAATCTTACTCTCATTTGATTTCCAGATTTGCATGAA	364
OY	1513	tatcaagctgtagcttaagcttaaaaaattggcctacaaagcaaatctatgttaggctta	1572
Db	365	TACCAAGCAGATGCTTTGGCCAAAGGATTTGGTTTACAGCAGATTTTGGCACAAAGCTTTA	424
OY	1573	atgatctacaacaataaaacctttccacacatgaatgtagatcctcttattctcagcat	1632
Db	425	ATTAACTTACAGATTTAAAGACTTGTGCCAATGATGTAGAACCACTTGTACTCGGCATAC	484
OY	1633	cataattcccatccaactctctagctgaaagatcgacgcgctctagactatglttagtgaag	1692
Db	485	CACACTCTTCACCCCAACTCTAGCCGAAGCTTAAACCGATTGGGGCTTGTAAGTCGCAAG	544
OY	1693	aagaaanaactatctatagat	1714
Db	545	AAGGCAGAGTGTGTAATTAAGT	566

RESULT	6			
LOCUS	A2930382			
DEFINITION	A2930382	517 bp	DNA	GSS
	474.dhz3f03.sl	Saccharomyces unisporus NRRL Y-1556 Saccharomyces unisporus genomic clone 474.dhz5f03.sl, DNA sequence.		

VERSION	AZ930382.1	GI:13501291
KEYWORDS	GSS.	
SOURCE	Saccharomyces unisporus	

ORGANISM Saccharomyces unisporus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE	AUTHORS
1 (bases 1 to 517)	Cliften, P. F., Hillier, L. W., Fulton, L., Graves, T., Miner, T., Gish

TITLE W.R., Waterston, R.H. and Johnston, M. Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
Journal of Molecular Biology (2001)

JOURNAL
UNPUBLISHED (2001)
COMMENT
Contact: Johnston M

Department of Genetics
Washington University Medical School
Box 8232, 4556 Scott Ave., St. Louis, MO 63110, USA
Tel.: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.

FEATURES	Location/Qualifiers
source	1. .517

```

/organism="Saccharomyces unisporus"
/strain="NRRL Y-1556 (CBS 396)"
/db_xref="taxon:27294"
/clone_1b="474.dh253f03.s1"
/clone_1lb="Saccharomyces unisporus NRRL Y-1556"
/note="Random genomic sequence"

```

Query Match	11.7%;	Score 213.6;	DB 13;	Length 517;
Best Local Similarity	63.4%;	Pred. No. 2.4e-26;		
Matches 327;	Conservative 0;	Mismatches 189;	Indels.: 0;	Gaps

Dd	2	TACAAAGTTTATATGCAAAACAACTTACCTCCACTTCTTGAAAGTGAAATTGACAAAGAA	61
Oy	514	actttcaataacaaggaactactcccggccaaggccaagttccactttcgglyac	573
Dd	62	ACTTTTGAANAATCGAAGAATATTCCAAAACGTAAGAATTAGTCTCATGCTCTGAT	121
Oy	574	gtctataaccaggcccaaaagctggttttcttcaataagcacccttcctccaaactcgg	633
Dd	122	ACCATCGGTTTACTCCAAAACATTGCTTTCATTAATAATGATATCTTACCAAAATTAATGG	181
Oy	634	caacatgcgcgtttcttcaatgaagcagtcctgcacgcagatccataatgctccac	693
Dd	182	AATTTTGGTCATATGATATGCGAAGCTGTCTTCCCTCGTAATATCAAGCTGTCTAT	241
Oy	694	gtcgcaagagttatagcttccttgggtctctctatccagttgtctaccttgglybaltg	753
Dd	242	GTTTGCTCAAAAGTTTGATGTTTATATTCACATATGCTAATTAATATACATGTTTGGCTTA	301
Oy	754	ccaactctactactagcatttgccttggtaaaaaaatttggttccaataatgacc	813
Dd	302	CCAAATTCGTATATATTCATATTCGTCTTAAACAAAAAATTTGGTTTAATAGTAACT	361
Oy	814	gtccaacatcgtatcacccgatcatgatcaagaagctcgaacttggcgtaatgctatggcgc	873
Dd	362	GTTAATATTAATGATTAATAGCATATTAAGGTAAATTTGTTAGGTTATGTGATTTGGTGT	421
Oy	874	ccaatcccttaaccgtttcccttaagaacttlygataaattccctactgaattcccttggtaac	933
Dd	422	CCAAATTCATTAATTAATCTTAATAAGATTTTATTAATTTGCAAACTAATTTCTTAATGCTAT	481
Oy	934	attatagctctcgttcgttcgtccaactctaaccc	969
Dd	482	AATTTGTTAATTTTCTCTGTTGTCGAAGTGTTTACC	517

DD 402 ALLIGILLICCAAGIGIAGCC JI/

RESULT	7
CNS06CBM/C	
LOCUS	
DEFINITION	
CNS06CBM	1116 bp
T7 end of clone AR06A003F08 of library AR06A from strain CBS 732 of <i>Zygosaccharomyces rouxii</i> , genomic survey sequence.	DNA GSS 14-JUN-2001

ACCESSION	AL392312
VERSION	AL392312.1
	GI:12141541

KEYWORDS	GSS.
SOURCE	Zygosaccharomyces rouxii

ORGANISM

Eukaryota; fungi; Ascomycota; sacchariomycotina; saccharomycetes, Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.

REFERENCE	AUTHORS
1 (bases 1 to 1116)	de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,

TITLE Wincker, P., Artiguenave, F. and Souciet, J. Genomic exploration of the hemiascomycetous yeasts. 8

Zygosaccharomyces rouxii
CBS 1444 (07/11) 50 ff. (2000)

JOURNAL
MEDLINE
FEB5 1991, 32-33 (2000)
20584718

REFERENCE	AUTHORS
2 (bases 1 to 1116)	Souciët, J. L., Aigle, J.

Bolotin-Fukuhara, M.,
de-Montigny, J. A. Dutoit

Malpertuy, A., Neugeglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Texada, F., Iorlano-Niocne, C., Wasolowski-Louvel, M., Wincker, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL
FEBs Lett. 487 (1), 3-12 (2000)
MEDLINE
20584711

REFERENCE 3 (bases 1 to 1116)

AUTHORS	Genoscope.
TITLE	Direct Submission

JOURNAL

COMMENT
This GSS is part of a random genomic sequencing program of thirteen

Db 560 CTTTCGCCCTACTACTGCTTTGTTGTTCTCCACATGCTGCTATGTAACGCCG 501
 Qy 980 ttcaagcttcacatcccaatgtaataagttacacccatcttgagaggtgaactga 1039
 Db 500 TTCTACTCTGATCATGCGCATGTTCAACAAGTTGAAGACCTTGAGAGCGCGAGCTCA 441
 Qy 1040 aaaaactatgaaagtgtgacgataagttggtgtccctcctagataagattttgtca 1099
 Db 440 AGAGCCGATCGGAGGCTCTTGCCAGAAAGCTGGAGCTTCCCTTTCCGACCTTTATGCA 381
 Qy 1100 ttgacgctcaaaaagatcttcacatcaaacgataatttcacaggttgccatcact 1159
 Db 380 TTGAGCGCTACCCGATCCGCTCAGTCACATGCTTCTTCCGCTTTCGCTTGAAG- 320
 Qy 1160 ccaagagaatgtttgttgacacacttaagtaaacagtaattctactgataaataagg 1219
 Db 321 -AAGCAGATTGCTCTGACGACACTGATGAGCAATGCTCCACACGAGATTGAGG 264
 Qy 1220 ctgtttggcccaatgaaatcgctacgtgcaaaaaaacacatcgtaataatgtgcatct 1279
 Db 263 CCATCTTGCGCCAGAGCTGGTCTACGTGAAGATGAACATATCTCTGACAGCGCTGCTG 204
 Qy 1280 ttatgaattgacacacttcctcatttctccctttaccagacatctacagaatacat 1339
 Db 203 CCGGCAACGCAACATTTCTGACTGACTGATTCCTGCTGCTGCTGCTGCTGCAACGACT 144
 Qy 1340 catttacaacacttgagcttttc 1365
 Db 143 CTTTTCACCTTMCCTGCGCTTCTTC 118

RESULT 9
 LOCUS B6176721 859 bp mRNA EST 06-FEB-2001
 DEFINITION 602313533f1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4419344 5',
 mRNA sequence.
 ACCESSION B6176721
 VERSION B6176721.1 GI:12683424
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Cetartihia; Homnidae; Homo.

REFERENCE 1 (bases 1 to 859)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Louis Staudt, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM10155 row: a column: 09
 High quality sequence stop: 492.
 Location/Qualifiers
 1. 859

FEATURES
 SOURCE
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4419344"
 /clone_id="NIH_MGC_85"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph. Vector: pCMV-SPORT6, site:1: Not;
 site:2: SalI. Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.867 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: This is a NIH_MGC Library."
 BASE COUNT 214 a 209 c 171 g 265 t

ORIGIN
 Query Match 5.9% Score 108.4; DB 11; Length 859;
 Best Local Similarity 54.6%; Pred. No. 7.3e-09;
 Matches 239; Conservative 0; Mismatches 196; Indels 3; Gaps 1;
 Qy 766 tatagccatttgcctcctgaaagaaatctggttccaataatgacccccaatgag 825
 Db 11 TATATACTTTTGATGATGAAGAAAAACATGCTTCATCAACAGACTTTGGGCTTTC 70
 Qy 826 atccacgatalgatcaagagctgacttggcgtatgtatgtgtggtgccaatccttac 885
 Db 71 ATGAAGATGCAATCAAGAAATTTGTTGTAAGTCAAGTATTTTGTCCGTGCTTCA 130
 Qy 886 ctgttccttaagatcttggataaatccctactgattcccttggtaatatgctc 945
 Db 131 CTTTACTTTACATTTATTAATAATGGGGGTACATTTTATTTATTAAGCCTGCTGTTG 190
 Qy 946 ttgttcgttgcacaaactttagccatgacacatctcagcttcacatgccaatgttt 1005
 Db 191 ACATTAGTTGTCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 250
 Qy 1006 aataagttacatccatgagagcgttgaactgaaataatctatgaaagtttgcgat 1065
 Db 251 GACAAATTCACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310
 Qy 1066 agagttgggtccctctatagaagaatttggtaattgacagcgtcacaagaatctctcat 1125
 Db 311 AGTATGACTTCTCTTACGACGAGGTGTATGCTGTGAGGAGTCTTAAACGCTCCAC 370
 Qy 1126 tcaacgcatatttccacaggttggcattccctcaccacaaagagaatgtttgtgacact 1185
 Db 371 AGCAATGCTTATTTTATGAGCTT---CTTCAAGAACAGAGCACTGTTTGTGTAAGCT 427
 Qy 1186 ttatgtaacagtaattct 1203
 Db 428 CTACTAGAAAGTACTCT 445

RESULT 10
 LOCUS AA965341 424 bp mRNA EST 31-JUL-1998
 DEFINITION e9b02a1.r1 Aspergillus nidulans 24hr asexual developmental and
 vegetative cDNA lambda zap library Emeritella nidulans cDNA clone
 e9b02a1 5', mRNA sequence.
 ACCESSION AA965341
 VERSION AA965341.1 GI:3139225
 KEYWORDS EST.
 SOURCE Emeritella nidulans.
 ORGANISM Emeritella nidulans.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; Emeritella.
 REFERENCE 1 (bases 1 to 424)
 AUTHORS Kuper, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
 Prade, R. and Roe, B.
 TITLE An Aspergillus nidulans EST Database
 JOURNAL Unpublished (1998)
 COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parlington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: T3
 High quality sequence stop: 351.
 Location/Qualifiers
 1. 424

FEATURES
 SOURCE
 /organism="Emeritella nidulans"
 /strain="FGSC A26"


```

Db 517 ACGGAGACCTAAAAAGACATTCAGATCTGGCAAGCAGTGTGAATTTCTTGGGCC 518
Oy 1088 agatttgcattgacggcgcacaaagatcttcattcaaacgcacatattcacaggtt 1147
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 517 AACTTTTGTCTCGAAGGTTCAAGCGTTCTGGCGCATAGTACGCTTACTTTACCGGAT 458
Oy 1148 tg 1149
    | |
Db 457 TG 456
    
```

Search completed: February 12, 2002, 12:01:41
 Job time: 6514 sec

THIS PAGE BLANK (23PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 11:09:13 : Search time 91.73 Seconds

(without alignments)
4505.858 Million cell updates/sec

Title: US-09-165-460a-1

Perfect score: 1825
Sequence: 1 accatactttttctatct.....aaataacagcttaaacatt 1825

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	7.5	686	4 US-08-998-416-929	Sequence 929, App
2	117.4	6.4	2968	3 US-09-022-699-1	Sequence 1, Appli
3	54.6	3.0	7218	1 US-08-232-463-14	Sequence 14, Appli
4	51.6	2.8	1641	1 US-08-300-903A-8	Sequence 8, Appli
5	49	2.7	1368	3 US-08-874-563-5	Sequence 5, Appli
6	49	2.7	1368	3 US-08-577-483-14	Sequence 14, Appli
7	48	2.6	8220	2 US-08-568-459A-11	Sequence 11, Appli
8	48	2.6	8220	2 US-08-487-826B-11	Sequence 11, Appli
9	48	2.6	19124	2 US-08-487-826B-13	Sequence 13, Appli
10	47.8	2.6	2035	3 US-09-022-699-3	Sequence 3, Appli
11	47.4	2.6	1016	1 US-08-109-391A-3	Sequence 3, Appli
12	47.4	2.6	1016	1 US-08-459-019A-3	Sequence 3, Appli
13	47.4	2.6	1016	1 US-08-460-428A-3	Sequence 3, Appli
14	47.4	2.6	1016	3 US-08-458-860A-3	Sequence 3, Appli
15	47.4	2.6	4253	3 US-08-577-483-7	Sequence 7, Appli
16	47	2.6	870	2 US-08-420-629-9	Sequence 9, Appli
17	46.8	2.6	4254	4 US-08-443-639-7	Sequence 7, Appli
18	46.8	2.6	80246	4 US-09-078-294-4	Sequence 4, Appli
19	46.2	2.5	1733	3 US-09-073-569-1	Sequence 1, Appli
20	46.2	2.5	6727	3 US-08-629-643A-5	Sequence 5, Appli
21	46.2	2.5	6727	4 US-09-280-799-1	Sequence 1, Appli
22	46.2	2.5	6727	4 US-09-155-884-5	Sequence 5, Appli
23	46.2	2.5	7218	1 US-08-232-463-14	Sequence 14, Appli
24	46	2.5	2082	2 US-08-785-310A-2	Sequence 2, Appli
25	45.8	2.5	2447	2 US-09-014-969-14	Sequence 14, Appli
26	45.8	2.5	144	1 US-08-702-344-26	Sequence 26, Appli
27	45.8	2.5	240	1 US-08-628-417-6	Sequence 6, Appli

28	45.6	2.5	1051	4 US-09-245-041-10	Sequence 10, Appli
29	45	2.5	1890	6 5312912-3	Patent No. 5312912
30	45	2.5	3072	4 US-09-522-217-55	Sequence 55, Appli
31	44	2.4	1117	4 US-08-247-373B-33	Sequence 33, Appli
32	44	2.4	8920	2 US-08-446-855A-1	Sequence 1, Appli
33	44	2.4	8920	4 US-09-150-741-1	Sequence 1, Appli
34	43.8	2.4	17949	4 US-09-087-465-3	Sequence 3, Appli
35	43.4	2.4	2551	1 US-08-482-279-70	Sequence 70, Appli
36	43.4	2.4	2551	2 US-08-482-279-70	Sequence 70, Appli
37	43.4	2.4	2551	2 US-08-342-268-70	Sequence 70, Appli
38	43.4	2.4	2551	3 US-09-015-868-70	Sequence 70, Appli
39	43.4	2.4	2551	4 US-09-397-386-70	Sequence 70, Appli
40	43.4	2.4	4718	3 US-08-936-135-9	Sequence 9, Appli
41	43.4	2.4	4733	3 US-08-936-135-11	Sequence 11, Appli
42	43.4	2.4	4769	3 US-08-936-135-13	Sequence 13, Appli
43	43.4	2.4	4784	3 US-08-936-135-15	Sequence 15, Appli
44	43.2	2.4	300	1 US-08-222-177A-45	Sequence 45, Appli
45	43.2	2.4	3481	4 US-08-965-729A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-998-416-929/c
Sequence 929, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippesen, Peter
APPLICANT: Pohmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 929:
SEQUENCE CHARACTERISTICS:
LENGTH: 686 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1577RP

US-08-998-416-929

Query Match 7.5%; Score 136; DB 4; Length 686;
 Best Local Similarity 52.1%; Pred. No. 6.1e-23;
 Matches 354; Conservative 0; Mismatches 320; Indels 6; Gaps 2;

OY 490 ctggaagaagaatgagatgaactttcaataatcaagaactaccgggccaag 549
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 675 ctggagggcggttatgtatgaagaacaaacaaagacggcctacgagggccaag 616
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 550 gccaaagtccatcttgcggtagcgtacaaacctagcccaaaagctagtttcaaca 609
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 615 TAACCTACCCGATGATGATAGATCTGGTCTCGGCACTAACTGCTATGATTAAG 556
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 610 taagacccttcccaaaatcgcacatgacgcttcttcttaagagcgtctgca 669
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 555 TAAGAGTGGCTGCGACGGATGTGGAACCTAGAGTGGCGTATGCTACGGC 499
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 670 gtcaaatcatalgtgtccacgtgcgacagattatgtcttctgtctctatcc 729
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 498 GCCATGCTGTCGTCAGATGATGACATCTGCGATGCTGCTATTTTATGCTTACTA 439
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 730 agttgtacacctgtgtgattgtccactctctactaagccatttgccttggaaga 789
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 438 CAGCTGAACGTGGTGGCAGGGGTTGTTGGAGACTTACTACACTTGTGCTCGAGGA 379
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 790 aaattgtttcaataaattgacgcgtccacatgatacaccgataatgacaagctcg 849
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 378 AAATTCGGGTTCACAAAGTCCACGCTCAACGTATGCTAACGGACGACGATAAAGTGT 319
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 850 acctggcgatgcatgtgtgcccacatcccttaccctgtcttccaaagcttgaataa 909
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 318 ATGATCTCTCCATGATTAATACCATCTCTGAGCCTATGCGCTTGAAGATATTGAAG 259
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 910 ttcctactgatttcccttggtagaatgattgctctgtgtgtgtgtgtgtgtgtgtgt 969
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 258 TTTCTGACGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 199
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 970 atgaaatcattccagcttctcattcattgacatgttaatttaaccatccatagagac 1029
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 198 ACTGCTCTTCAAGCCAGCTTT--TTACAGCTTTATTCAACAGCTTACCCGTTGAGAGC 142
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1030 gttgacatgaaaaaaatctatgaaagtgtgcccagtagagtggtgtccctctagataa 1089
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 141 GGGAGGCTGAAGAGCGTTCATTTGCAAAATTAAGCAAAAGGTAACCTCCCATGAGACAA 82
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1090 attttgtcatgacggtcctcaaaagatcttctcaatcaaacgcatattcaaggttg 1149
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 81 ATCTACCTTTTCGAGGCGCTCGCGCTCTCGGCACTCAAAATGCGTATTTCACCGGCTA 22
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1150 ccattccactccaagaagaat 1169
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 21 CCGTTTTTCTTAAACGAT 2
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2

US-09-022-699-1
 Sequence 1, Application US/09022699
 Patent No. 6060277

GENERAL INFORMATION:
 APPLICANT: KIKLY, KRISTINE
 APPLICANT: SOUTHAN, CHRISTOPHER
 APPLICANT: KNAB, ANNE
 TITLE OF INVENTION: Human AFCL
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: RAYNER & PRESTIA
 STREET: P. O. BOX 980
 CITY: VALLEY FORGE
 STATE: PA
 COUNTRY: USA
 ZIP: 19482

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/022,699
 FILING DATE: 12-FEB-1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 97304440.7
 FILING DATE: 12-JUN-97
 ATTORNEY/AGENT INFORMATION:
 NAME: PRESTIA, PAUL F
 REGISTRATION NUMBER: 23,031
 REFERENCE/DOCKET NUMBER: GH-70380
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-407-0700
 TELEFAX: 610-407-0701
 TELEX: 846169
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2968 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-09-022-699-1

Query Match 6.4%; Score 117.4; DB 3; Length 2968;
 Best Local Similarity 50.5%; Pred. No. 1.8e-18;
 Matches 370; Conservative 0; Mismatches 351; Indels 12; Gaps 3;

OY 471 aacaaagtgtccacctgtgtcgtggaagcgaatgatagtgaactttcataaatcaag 530
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 177 AACCTATGTACCCGAGGATTAGACACATGATGATGATGATGATGATGATGATG 236
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 531 gaactactccggcgccagagccaaagtctcatttgcgtgtgagtcataaccctagcca 590
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 237 ACTCTATCACTGATTAAGCACTTTCAGCTTTCGTCAGAGCTCTTTCAGAGACTGA 296
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 591 aaagctagtttcaataaatacgaaccttccctaaatcttgcacatgacgtgttctt 650
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 297 AGGCACTCTTAATTTCTTCTTTGGAGGAATACCTTATCTGAGAGACTTTCGACGCTT 356
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 651 attgaatgcagtcctgcagtcagatctcaatgattgtctccactgtgcacagattatg 710
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 357 CT--GTGGTTATGCTGGCTTTGGACCAAGAAATGAGATCACTCCCTGGTGTCTG 414
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 711 ctcttgggtctctatccagtttgttaccctgtgtgattgtgacactctctactatag 770
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 415 CTGTTGGCTTACACT-----TTTCAGTGCATTGCTGTTGGCATGAGACTTTTATTA 467
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 771 coattttgtctggaagaaaaatttgttcaataaatgaaacgttccaactatgagatcac 830
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 468 TACTTTGTGATAGAGAAAAAACAATGCTTCAATCAACAGACTTTGGGCTTCTTCATGAG 527
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 831 cgatatgataagagctcgaacttggcgtagctatgtgtgtgtgtgtgtgtgtgtgtgt 890
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 528 AGATCAATCAAGAAATTTGTTGTGACACAGTATTTGTTGCTGCTGCTTCACTCTCT 587
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 891 ccttaagaattttgataatccccaactgatttcttctgtgacattatggtctcttctgt 950
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 588 ACTTTACATTTATTAATTTGGGGGTGACTATTTTATTTATGCTGCTGCTGTTACATT 647
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 951 cgtttccaatcttagacatgacatcaatccagcttccatgacatgacatgacatgacat 1010
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 648 AGTTGTGCTCGTTCTTGTCACAACTATGCTGATTAATTTGGCCCTTATTTAGACAA 707
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 1011 gtcaactcatttggagagcgtgtgaactgaaaaaactatgtgaagtgttggccgataagct 1070
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 708 ATTACACCTCTGCTGAGGAGAAAGCTTAAAGAGAAATGGAATGAGCAAGAGATAT 767
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Query Match	2.7%	Score 49;	DB 3;	Length 1368;
Best Local Similarity	64.6%	Pred. No. 0.01;		
Matches	73;	Conservative	0;	Mismatches 40;
			Indels	0;
			Gaps	0;

GENERAL INFORMATION:

```

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
City: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH21.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 8220 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEITICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-11

Query Match      2.6%   Score 48: DB 2: Length 8220:
Best Local Similarity 59.6%   Pred. No. 0.027:
Matches 81: Conservative 0: Mismatches 55: Indels 0: Gaps 0:

QY    20 ttcaacagcgaagaaggcttacacacacacacatatacatatcacatatacaata 79
      ||| || | | | | | | | | | | | | | | | | | | | | | | | | |
Db    8019 TACATATACAGAAATCATATACATACACATATATATATTCCAAAAACATACATAA 7960

QY    80 tacatatgttaaaccttgttatattcccttataaccaaaaggagcaattaacttt 139
      ||||| || | | | | | | | | | | | | | | | | | | | | | | |
Db    7959 CATATATATATATATATATATATTCACAACATATACACAATACATATTAACACAT 7900

QY    140 cccctctttcttagct 155
      || | | | | | | |
Db    7899 CCACATACATCCACAT 7884

RESULT          9
US-08-487-826B-13/c
Sequence 13: Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitluis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor

```

RESULT 11
 US-08-109-391A-3/C
 Sequence 3, Application US/08109391A
 Patent No. 5639876
 GENERAL INFORMATION:
 APPLICANT: Trilip, Cynthia A.
 APPLICANT: Frank, Glenn R.
 APPLICANT: Grilvee, Robert B.
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NOVEL
 TITLE OF INVENTION: PARASITIC HELMINTH PROTEINS
 NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sheridan Ross & Mcintosh
 STREET: 1700 Lincoln St., Suite 3500
 CITY: Denver
 STATE: CO
 COUNTRY: U.S.A.
 ZIP: 80203
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/109,391A
 FILING DATE: 19-AUG-1993
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Connell, Gary J.
 REGISTRATION NUMBER: 32,020
 REFERENCE/DOCKET NUMBER: 2618-13
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 303/863-9700

PatentNo. 591237
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838

THIS PAGE BLANK (23PT0)

[illegible]

XX (RESC) UNITV CALIFORNIA.
PA Ashby MN, Boyarchuk VL, Rine JD;
XX WPI: 1998-145621/13.
XX N-PSDB: AAV17601.
DR
XX
XX Vector comprising nucleic acid coding for Afcip and Rceip proteins
PT involved in protein prenylation - useful for screening for
PT inhibitors of proteins, especially of mutated ras proteins involved
PT in cancerous conditions
XX
XX Claim 15; Page 50-52; 59pp: English.
PS
XX
XX This polypeptide comprises yeast Afcip (a-factor converting
CC enzyme), a novel zinc metalloprotease that mediates the removal of
CC AAX tripeptide from CAAX proteins following prenylation. Its amino
CC acid sequence was deduced from the isolated yeast AfcI gene (see
CC AAV17601). Genetic knockout of AfcI resulted in incomplete reduction
CC of a-factor production, thereby exposing the presence of a second
CC CAAX protease. This was designated Rceip (Ras and a-factor
CC converting enzyme, see AAV48302). Vectors comprising the AfcI or RceI
CC gene (see AAV17602), polypeptides encoded by such vectors and
CC recombinant cells transfected with the vectors are claimed. The
CC novel Afcip and Rceip proteins can be used to screen for inhibitors
CC of prenylation which can be used in the treatment of Ras dependent
CC cancers such as colorectal and exocrine pancreatic carcinomas or
CC myeloid leukaemia. A heat shock assay for assessing ras activity
CC is also provided.
XX
XX Sequence 453 AA:
SO

Query Match 98.9%; Score 2317; DB 19; Length 453;
Best Local Similarity 99.3%; Pred. No. 2.1e-238;
Matches 449; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MFDLKTLIDHPNIPWKLITSGSIAOFSESYLYROYOKLSEKLPVLEDEIDDETFH 60
DB 1 mfdltklidhpnipwklitsgsifsfeslylryqyklsctklpvledeldetfhn 60
QY 61 KSRNRSRAKAKFTGCDVYNLAOKLVFTKDYLPFKIMHANSLSLAVALPVRHMVSTVAO 120
DB 61 ksrnrsrakakftsgcdvynlaoklvftkydlypfkimhanslslavalpvrhmstvao 120
QY 121 SLCPGLSLSLSTLVLDPLSYSHFVLEBEKFGFNKLTVOLMTDMIKSTLTAAYAGPIL 180
DB 121 slcpglslslstlvldplsyshfvlbeekfgfnkltvolmtdmikstltaayagpil 180
QY 181 YLFLEKIDFKPPTDFLWYIMVFLVVOQLAMTIPVIMPMENKFTPLEDGELKKSIESLA 240
DB 181 ylflekidfkpptdflywimvflvvoqlamtppvimpmenkftpledgelkksiesla 240
QY 241 DRVGPFLDKIVIVOSKSSSHSNATFTGLPFTSKRIYVFDPLVNSNSTDETTAVLAHEIG 300
DB 241 drvpgfldkivivoskssshsnatftglpftskriyvfdplvnsnstdettaavlheig 300
QY 301 HMOKHHIYNNVYIFSQLHTFLIFSLFTSYRNTSFYNTFGFLEKSTGSEFVDPVITKEPPI 360
DB 301 hmokehhiyynnvyifsqlhtflifslftsyryn timerntsfyntfgflekstgsefvdpvitkeppi 360
QY 361 IIGFMFLNDLLTPLECAMQFVMSLSTRTHEYQADAVAKKLYKONIKRALIDLQIKNST 420
DB 361 iigfmflndlltplecamqfvm slstrtheyqadavakklykonikralidldqiknst 420
QY 421 MNVDPVLYSSYHSHPTLAERSTALDYSEKK 452
DB 421 mnvdpvlyssyhshtlaersta ldysekk 452

RESULT 2
AAG15512

ID AAG15512 standard; Protein; 458 AA.
XX
XX AAG15512;
AC
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 15793.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
KM
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.

PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150366.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140891.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-01

PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156448.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160960.
PR	22-OCT-1999;	99US-0160961.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

36.5%; Score 856; DB 21; Length 456;

Best Local Similarity 41.2%; Pred. No. 17e-82;

Matches 185; Conservative 72; Mismatches 167; Indels 24; Gaps

QY	6	TILDHPIPWKLIIISGEISIAOFSESYITRYOYKLSETKLIPVLEDEIDDETFHKSRY	65
DB	30	tlfsamaprlnevtvgflmvyllfetyldlrqltalxlpclprklvgvisqekfeksray	89
QY	66	SRRAKFSIFEDDVYNLQKLVFIYDLEPRKIMHMAVSLNNAVLV-VEFHMAVSIVAOSLCF	124
DB	90	sldksylfivhefvtlmdsaillffglpwfwkms----gavprlridopenellnlstf	145
QY	125	LGLLSLSTLVLDPLSVYSHFVLEEEKGFENKLFVQALMITMIXSLTAAVAGPIYFL	184
DB	146	lagymtvgqldldpfslystvisrningfoktlfmflrmlkgtflsvllgppivaal	205
QY	185	KIEDEPFDLWYIMVLEFVQIILAMTIIPVFIWPFNKFTPLEDEGLKSIIESLADRVG	244

PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156448.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160960.
PR	22-OCT-1999;	99US-0160961.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

36.5%; Score 856; DB 21; Length 456;

Best Local Similarity 41.2%; Pred. No. 17e-82;

Matches 185; Conservative 72; Mismatches 167; Indels 24; Gaps

QY	6	TILDHPIWPKLIISGESIAOFSESYITYROYKLSETKLIPVLEDEIDDETFHKSRY	65
DB	30	tlfsamaprlnevtvgflmvyllfetyldlrqltalxlpclprklvgvisqekfeksray	89
QY	66	SRRAKFEISFEDDVYNLQKLVFIYDLEPRKIMHMAVSLNNAVLV-VEFHMAVSIVAOSLCF	124
DB	90	sldksylfivhefrclmdsaillffglpwfwkms----gavprlridopenellnlstf	145
QY	125	LGLSLSTLVLDPLSVYSHFVLEEEKGFENKLFVQMLITMIMKSLTAAVAGPIYFL	184
DB	146	lagymtvgqldldpfslystvisrningfoktlfmflimlkgftslvllgppivaall	205
QY	185	KIEDEKPTDLWYIMVLEFVQILAMTIIPVFIWPFNKFTPLEDEGLKSIIESLADRVG	244

Db 206 flvkggpylalyafmfllslvmntlypvliaplfnkflrplpdgdlrekleklaasslk 265
Qy 245 FPLDKIFVDSGRSSHSNAFTGPEFTSKRIVLEFDITVNS-NSDDEITAVLAHEIGHMO 303
Db 266 flklklfvdgsttrsshaymygf-fknktivlydtliqgcknedelayavlahelghmk 324
Qy 304 KHHIYMWVFSOLHFLFLFSFTSYRMTSPYNTGFFLEKSTGSEVDVITKEPPIIG 363
Db 325 lnhctysflavqllafllgfygylvrnstldfrsfqfctg-----pvllg 369
Qy 364 FMLFRLDLPLECAMQFVMSLISRTHVQADAVAKKLCQKONLCRALIDLOIKNSTMWV 423
Db 370 lllfhtvlpqlhivstfgnlvstrafetqadafvklyakdlrpalvklgeenlsamnt 429
Qy 424 DPLVSSYHSHPTELAERSTALDYSEKKK 452
Db 430 dplysaeyhshpplverlraidd-gecdk 456

RESULT 3
AAG15513
ID AAG15513 standard; Protein: 424 AA.
XX AAG15513;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 15794.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN
XX
PD
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR	03-AUG-1999	9905-0147303
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147312
PR	05-AUG-1999	9905-0147360
PR	06-AUG-1999	9905-0147305
PR	06-AUG-1999	9905-0147316
PR	09-AUG-1999	9905-0147493
PR	09-AUG-1999	9905-0147935
PR	10-AUG-1999	9905-0148171
PR	11-AUG-1999	9905-0148319
PR	12-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148565
PR	16-AUG-1999	9905-0149368
PR	17-AUG-1999	9905-0149191
PR	18-AUG-1999	9905-0149426
PR	20-AUG-1999	9905-0149722
PR	20-AUG-1999	9905-0149929
PR	23-AUG-1999	9905-0149902
PR	23-AUG-1999	9905-0149930
PR	25-AUG-1999	9905-0150566
PR	26-AUG-1999	9905-0150884
PR	27-AUG-1999	9905-0151065
PR	27-AUG-1999	9905-0151060
PR	30-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151438
PR	01-SEP-1999	9905-0151930
PR	07-SEP-1999	9905-0152633
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154709
PR	22-SEP-1999	9905-0155139
PR	23-SEP-1999	9905-0155488
PR	24-SEP-1999	9905-0155659
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158032
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159331
PR	14-OCT-1999	9905-0159637
PR	18-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161350
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161922

[illegible]

PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 17-JUN-1999; 990S-0139453.
PR 18-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 26-JUL-1999; 990S-0145224.
PR 27-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145918.
PR 28-JUL-1999; 990S-0145911.
PR 02-AUG-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.

Query Match	36.2%;	Score 848;	DB 21;	length 419;
Best Local Similarity	41.7%;	Pred. NO. 1.1e-81;		
Matches 182;	Conservative 71;	Mismatches 159;	Indels 24;	Gaps 6

KM	Inhibitor; enzyme; degradation; Ras; diagnosis; cancer.
XX	Homo sapiens.
OS	
PN	W0935275-A1.
XX	
PD	15-JUL-1999.
XX	
PF	08-JAN-1999; 99WO-ES00001.
XX	
PR	08-JAN-1998; 98ES-0000016.
XX	
PA	(UYOV-) UNIV OVIEDO.
PI	(F0JY) FUJI YAKUHIN KOGYO KK.
PI	Perez Freije JM, Velasco Cotarejo G, Martin Pendas A, Blay Albers P;
DR	Balbin Fetechos M, Lopez Olin C;
XX	WPI, 1999-430402/36.
DR	N-PDSB: AAX85493.
PT	New human sequences encoding enzymes that degrade prenylated proteins,
PT	useful in diagnosis and treatment of cancers -
XX	
PS	Disclosure; Page 15-18; 28pp; Spanish.
CC	
CC	This sequence represents the human farnesylated-protein converting
CC	enzyme 1 (Face-1) which has homology to the Saccharomyces cerevisiae
CC	Afc-1 protein. Afc-1 is an enzyme involved in the proteolytic processing
CC	of prenylated proteins and is putatively a metalloprotease for
CC	processing factor-a. This protein and the Face-2 protein (AAV26897) can
CC	be used to design specific inhibitors of enzymes involved in degradation
CC	of prenylated proteins, particularly Ras and related proteins and for
CC	treating disease mediated by Ras and related proteins and/or the genes
CC	encoding them, specifically for diagnosis and treatment of cancers.
XX	
SQ	Sequence 475 AA;
Query Match	29.8%; Score 698.5; DB 20; Length 475;
Best Local Similarity	35.6%; Pred. No. 1.2e-65;
Matches 165; Conservative	75; Mismatches 165; Indels 61; Gaps
YY	
QY	22 FSIAQSFSESYLYTRQ--YOKLSEFKLPVLDEIDETFFKSNYSRAKAFFISGDV 78
DB	II : :::: I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
DB	26 fswtvlywetfiagqrriyk--ttbhvppeigjmdsetfeksrlgldkfstfsfwsgl 83
QY	79 YNLAOKLVFIKYDLEFPKIMHMAVSLNAVLPVRFMVSTVAOSLCFLGLSSLSLVLDL 138
DB	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
DB	84 yseegctllilffgiiprylwrisgrfcgy---agfpzeyltqstivflillatfsalcglp 140
QY	139 LSYSHVFLEERKEFGNKTLTLOULTMDMLKSLTLAAIGSPIYLPLTKTFDKRPDFELWYI 198
DB	II : : : : : II : : : : : III : : : : : IV : : : : : V : : : : : VI : : : : :
DB	141 wsllyntftvieekhfngqtligifmndaikfkvtvccllprrssillylikiggdyffilya 200
QY	199 MYPLEVVOLIAMTIIPVEIMPMPFNKFTPEDOELKKSIESLADRYGFPDLKIFVDGSKR 258
DB :
DB	201 wlftlvavslvltviadyiapldkfplpegklkeelevmaksidfpltkvyveggkr 260
QY	259 SSHSAVTGTGLPFSKRILVEDTLVNSN----- 287
DB	IIIIIIIIII : : : : : IIIIIIIIIII : : : : : IIIIIIIIIII : : : : :
DB	261 sshsnayfygf-fknkrivilfdtlileeysvlnkdigedsgmeprneeegnseeiakakvn 319
QY	288 -----DPEITAVLAHEIGHWOKNIIVMMVIFSOLHTFLIFSLFSIRNNSFYNTGGPF 341
DB	::: : : : : :::: :
DB	320 kkggcckneevlaavignheghwklghtvnnllisgmnsficflitavllgrkeliaafgy 379
QY	342 LEKSTGSFVDEVITKEPFIITIGFM-LFNDLLTPLECAQFWSLSTRTHYEOADAYAKL 400
DB	I :
DB	380 dsq-----ptliglliliffsfisynevfisfclyttsrfffetgaadfaakl 425
QY	401 GYKNLCRALIDLQIKNLSTMNVDPLYSYHYSHPTJAERSTAL 444
DB	I : : : : : II : : : : : III : : : : : IV : : : : : V : : : : : VI : : : :

Db 426 gkakdlysaliklnkdnlgfpvswlfsmwhysbpllerlqal 469

RESULT 6

AAW87587 standard; Protein; 474 AA.

AAW87587;

12-MAR-1999 (first entry)

Human metalloprotease AFCL.

human AFCL metalloprotease; AFCL; haFCL; vaccine; cancer; inflammation; autoimmunity; allergy; asthma; rheumatoid arthritis; osteoarthritis; neurological abnormalities; septic shock; sepsis; stroke; osteoporosis; ischemia reperfusion injury; cardiovascular disease; kidney disease; liver disease; myocardial infarction; hypertension; AIDS; hypotension; haematological abnormality; aplastic anaemia; male pattern baldness; infection.

Homo sapiens.

EP887416-A2.

30-DEC-1998.

18-JUN-1998; 98EP-0304827.

12-FEB-1998; 98US-0022699.

24-JUN-1997; 97EP-0304440.

(SMIK) SMITHKLINE BEECHAM CORP.

(SMIK) SMITHKLINE BEECHAM PLC.

Kikly K, Knab A, Southan C;

WPI: 1999-047883/05.

N-PSDB: AAV83752.

Claim 4: Page 14; 22pp; English.

The present sequence represents a human AFCL metalloprotease (haFCL) polypeptide. haFCL polypeptides and polynucleotides are useful for diagnosing diseases related to over or underexpression of haFCL protein by identifying mutations in the haFCL gene, and/or analysing for the presence or amount of expressed polypeptide. haFCL polypeptides can be used to screen for agonists and antagonists by measuring the binding to polypeptide, and observing the stimulation or inhibition of polypeptide function. haFCL polypeptides can be administered directly or as a vaccine to inoculate against disease. Diseases diagnosed, prevented and treated include: cancer, inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis, neurological abnormalities, septic shock, sepsis, stroke, osteoporosis, osteoarthritis, ischemia reperfusion injury, cardiovascular disease, kidney and liver disease, myocardial infarction, hypo/hypertension, AIDS, haematological abnormalities, aplastic anaemia, male pattern baldness, and bacterial, fungal, protozoan and viral infections. The haFCL polypeptide is also useful for mapping the gene to a chromosome, allowing gene inheritance to be studied through linkage analysis.

Sequence 474 AA:

Query Match 27.9%; Score 653; DB 20: Length 474;

Best Local Similarity 34.7%; Pred. No. 8.3e-61;

Matches 161; Conservative 73; Mismatches 168; Indels 62; Gaps 8;

QY 22 FSIADFESYLYTYRQ--YOKLSETKLPVLEDEIDETPHKSNYSRAKAFSIFGDV 78

Db 26 fswtvytwetflaqrgrlyk--ctthvppegqlmddetfeksrllygldstcfswgl 83

QY 79 YNLAOKLVFIKYDLPFKTIHMAVSLNANLVPVRFHMVSTVAOSLCEPLDLSSTLVDP 138

Db 84 ysetegtlilllfglpylwrtsgrfcgy---agfgeyelqstlvtvlllatlfsalgjp 140

QY 139 LSYSHFVLEKEFGFNKLTVOIMIDMIKSLTLAVAGPILYLFKFDKFRPTPLMYI 198

Db 141 wslntvtyeehngtqglgfmtdalkktvtqclllpvsslllylkigdyffilya 200

QY 199 WVELFVQIILAMTIPVFMFENKFTPLEDEGLKKSIESLADRVGFPLDKITFYDGSKR 258

Db 201 wlfetvelvltvlyadylapfdkftlpgsklleelevmaksdldlktkvyvegskr 260

QY 259 SSHSNAYRTGLPFPSKRIVLEDTLVNSNS----- 287

Db 261 shsnayfygffkxkrlvldclleesvlnkddqedsqmeprneegnselkavkn 319

QY 288 -----TDEITAVLAHEIGHQKNHIVMVFISQHTFLISLFTSIRNHSFYTFGFF 341

Db 320 kkgckneevlavlgheighw-kwdlgsklslareffvflfavllgrkelaaigfy 378

QY 342 LEKSTGSEFVDVITKEPFIIGFM-LFNDLTPLECAMQFVMSLISRTHEXQADAVAKL 400

Db 379 dsq-----pllglllftgfispynevsfclvtlsrrfegqdaefakl 424

QY 401 GYKONLCRALDLOIKNLSTMNVDPLVSSYHSHTLAERSTAL 444

Db 425 gkakdlysaliklnkdnlgfpvswlfsmwhysbpllerlqal 468

RESULT 7

AAW8297 standard; Protein; 340 AA.

AAW8297;

14-MAR-2001 (first entry)

Lung cancer associated polypeptide sequence SEQ ID 635.

Human; lung cancer associated protein; neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active; vulnerrary; gastrointestinal; nephrotropic; antiinfective; gynecological; antibacterial; diagnosis; neural disorder; immune disorder; reproductive; proliferative disorder; wound healing; infectious disease.

Homo sapiens.

WO200055180-A2.

21-SEP-2000.

08-MAR-2000; 2000WO-US05918.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

(ROSE/) ROSEN C A.

Ruben SM;

WPI: 2000-587514/55.

N-PSDB: AAF18173.

Lung cancer associated gene sequences, referred to as lung cancer antigens, useful for treatment, prevention, and diagnosis of disorders such as lung cancer -

Claim 11: Page 1132-1134; 1425pp; English.

Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer


```

RESULT 9
AAB46348
ID AAB46348 standard; Protein: 414 AA.
XX
AC AAB46348:
XX
DT 05-APR-2001 (first entry)
XX
DE H. pylori HPS187 protein.
XX
KM Microbial infection; antibacterial; Helicobacter pylori infection;
KM vaccine; screening.
XX
OS Helicobacter pylori.
XX
PN WO200073502-A2.
XX
PD 07-DEC-2000.
XX
PF 31-MAY-2000; 2000MO-EP05024.
XX
PR 31-MAY-1999; 99DE-1024965.
PR 17-JUN-1999; 99DE-1027740.
PR 21-JUL-1999; 99DE-1034029.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (CREA-) CREATOGEN GMBH.
XX
PI Apfel H, Fuchs TM, Glibbs CP, Hueck CJ, Meyer TF;
XX
DR WPI: 2001-049948/06.
DR N-PSDB; AAF25625.
XX
PT Preparing an agent for diagnosis or control of microbial infection,
PT useful particularly against Helicobacter, based on identification of
PT essential genes in defective mutants.
XX
PS Claim 37; Page 356-357; 366pp; German.
XX
CC This invention describes a novel preparation of an agent (A) for
CC detection, prevention and/or treatment of microbial infection by:
CC (i) identifying essential genes (I) and corresponding polypeptides
CC (II); (ii) identifying compounds that are directed against (II) and
CC inactivate the microbe; (iii) testing these for suitability for use; and
CC (iv) formulating selected (A). Identifying essential genes (I) comprises
CC preparation of gene-deficient microorganisms by conditional antisense
CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
CC then determining viability and/or survival of the deficient organisms.
CC The products of the invention have antibacterial activity. (A) (which may
CC be a nucleic acid (ia), vector or host cell containing (ia), derived
CC polypeptide (Iia), or fragments, (Iia)-specific antibodies or their
CC fragments or an inhibitor of (Iia)) are particularly used for diagnosis,
CC treatment or prevention of infection by Helicobacter pylori. Particularly
CC (Iia) and (Iia) are used in DNA, subunit or live vaccines. The method
CC identifies essential genes, including those that have homologs in other
CC species, so identified (A) should have a broad spectrum of activity. Many
CC gene-deficient cells can be screened quickly, in an automated process.
CC and the identified genes can be used for screening without purification.
XX
SO Sequence 414 AA:

```

Query Match 19.4%; Score 454; DB 22; Length 414;
 Best Local Similarity 30.7%; Pred. No. 1.2e-39;
 Matches 129; Conservative 84; Mismatches 151; Indels 56; Gaps 17;

```

OY 39 OKLSEKLPVLEDEIDETFRKSNYSRAKAFSIFGVDVYLAOKLVFIKYDLPKIMH 98
DB 42 qlce---kpyl---lpqkdyeeagnyalrkmqlslsgld---gllfagv-vffglth 91
OY 99 M-AVSLNAVLPVRFHNVSTVAVOSLCLGLSLSTLVDPLSYSHFVEEKFGPKL 156
DB 92 ledlthyn-lp---etlgyivfallflaigsyia---lpisyttcmhdikelfgfskv 142

```

```

OY 157 TWQLMTDMIKSLTLAVAGCPILYLFLKIFDKFPDPLMYIMF--LFVVOILAMTIIP 214
DB 143 slsllfkdfkglstlslsvglilllyclimliehveh---welslffvfvfmlanllyp 199
OY 215 VFIMPENKFPLEDEGELKKSIESLADRYGFPDLKIFVIDGSRSSHNAVPTGLPFSK 274
DB 200 k-laglfngfuplnnrldiesqlegmmdkvgfksqglfymdaskrgrlnayfgyig-knk 257
OY 275 RLVFDFLVNSNSTDEITAVLAHEIGHWOKNHIYVMVIFSQLHTLFLSFTS---IYR 330
DB 258 rvlfdtliskvgtegliallqhelghfknkollslsgimgliallalahlpplvfe 317
OY 331 NTSFYNTGFELEKSTGSPVDPIVIRKEPFIIGFW-LTNDLITPLECAMQFMSLSIRH 389
DB 318 -----gfnvsglpsall-----aillflpvlstfyampI-----lgtfsfkn 354
OY 390 EXQADAVAKKLGKYNOLCRALIDLOIKNLSTWNVPLSSYSHYSHTPLAERSTALDYSE 449
DB 355 eynackfgaslskskevlakalvsivsenkafpysphfvlhftphlplerlkaldyle 414

```

RESULT 10
 AAB95578
 ID AAB95578 standard; Protein: 235 AA.
 XX
 AC AAB95578;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:18234.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 18234; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of

PD 07-DEC-2000.
 XX 31-MAY-2000; 2000WO-EP05024.
 XX 31-MAY-1999; 99DE-1024965.
 PR 17-JUN-1999; 99DE-1027740.
 PR 21-JUL-1999; 99DE-1034029.
 XX (PLAC) MAX PLANCK GES. FORDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.
 XX
 PI Apfel H, Fuchs TM, Glbbs CP, Hueck CJ, Meyer TF;
 XX WPI: 2001-049948/06.
 DR N-PSDB: AAF25691.
 XX
 PT Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 PT essential genes in defective mutants -
 XX
 PS Claim 37; Figure 15; 366pp; German.
 XX
 CC This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
 CC then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) (which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 CC fragments or an inhibitor of (IIa)) are particularly used for diagnosis,
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process,
 CC and the identified genes can be used for screening without purification.
 XX
 SQ Sequence 357 AA:

Query Match 16.6%; Score 389; DB 22; Length 357;
 Best Local Similarity 30.3%; Pred. No. 8e-33;
 Matches 115; Conservative 78; Mismatches 131; Indels 56; Gaps 17;

39 OKLSETKLPVLEDEIDETFKHSRNYSAKAKFSIFGVDVYMAQKLVETIKYDLPPKIWH 98
 25 qlcice---kpyl---lpqkdyeeagayalrkmgslslsgld---giltagv-vffgltgh 74
 99 M-AVSLNAVLVPRFHMVSTVAQSLCFLGLSLSTLVLDPLSYSHFLEKEKGFNKL 156
 75 ledlchlyln--lp---etglyivfallflaqlsyla---lpisyytlmhkdeqfskv 125
 157 TVQWMTMDIKSLTAVAGPILVLEFLKIFPKFPIDFLMYIMV--LFVQVILAMTTP 214
 126 slstlfdtkfgslstlsvgllyllylmlelhveh---wealsffvfvfmlanllyp 182
 215 VTIMPMFNKFTPLEDEGLKSLIESLADRVGFPLDKIFIVDSKRSRSHNAVYGPFTSK 274
 183 k-laqflngfcpimnrldesqlesmdkvkfseglfmdaaskrgrlnayfg919-knk 240
 275 RIVLEDTLVNSNSTDEITAVLAHEIGHMOKNHIYVMVIFSOLHFLISLFTS---IVR 330
 241 rylvldclskvgtelgiallghelghfnkdlkslgmgglalvalahlpriyde 300
 331 NTSFNTGCFLEKSTGSDVPVITKEPPIIIGFM-LFNDLITPLECAMQFVMSLISRTH 389
 301 -----gfnvsqfpasli-----tlllflpvlsfyampj-----igfssfn 337

QY 390 EYQADAYAKKLGKONLCRA 409
 DB 338 eynackdfgslskvylaka 357

RESULT 13
 ID AAY05501 standard; Protein; 426 AA.
 XX AAY05501;
 AC AAY05501;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE Bacillus subtilis metalloprotease YHN.
 XX
 KW Metalloprotease; protease; YHN; textile; detergent; feedstuff;
 KW animal feed; host expression system.
 XX
 OS Bacillus subtilis.
 XX
 FH Key Location/Qualifiers
 FT Region 277..281
 FT /note="HEXXH metalloprotease motif"
 XX
 XX WO9914341-A2.
 XX 25-MAR-1999.
 XX 08-SEP-1998; 98WO-US18677.
 XX 15-SEP-1997; 97GB-0019637.
 XX (GEMV) GENENCOR INT BV.
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Estell DA;
 DR WPI: 1999-244036/20.
 DR N-PSDB: AAX25097.
 XX
 PT New metalloproteases derived from gram-positive microorganisms
 PT
 PS Claim 12; Fig 1A-E; 54pp; English.
 XX
 CC This protein is a novel metalloprotease (MP), designated YHN, of
 CC Bacillus subtilis strain 1-168. YHN has overall amino acid
 CC homology to Saccharomyces cerevisiae Stre24, and its gene (see
 CC AAX25097) is found immediately downstream of the major alkaline
 CC protease (apre) gene putative transcriptional terminator,
 CC suggesting a role in regulating and/or processing the major
 CC alkaline protease. The invention relates to the identification of
 CC novel MPs in gram-positive microorganisms, especially B. subtilis,
 CC Bacillus stercorophilus, Bacillus licheniformis and
 CC Bacillus amyloliquefaciens. The invention also provides methods
 CC for the production of MP in host cells as well as for the
 CC production of heterologous proteins in a host cell having a
 CC mutation or deletion of part or all of the MP. Preferred
 CC heterologous proteins are hormones, enzymes, growth factors and
 CC cytokines. Gram positive MPs have applications in the textile
 CC industry, in cleaning compositions (claimed), and in animal feed.
 CC They can be used in bar or liquid soap applications, dish care
 CC products, contact lens cleaning solutions and products, peptide
 CC hydrolysis, and waste treatment.
 XX
 SQ Sequence 426 AA:

Query Match 14.0%; Score 328; DB 20; Length 426;
 Best Local Similarity 26.9%; Pred. No. 3.4e-26;
 Matches 108; Conservative 76; Mismatches 152; Indels 66; Gaps 14;

81 LAOKLVETIKYDLF-----PKIMHMAVSLD-----NAVLPVRHMVSTVAQSLCF 124


```

Db 53 vaegynvknflffigvpldwflffvllvsgvskkikkwleaaavfrf-----lqtv9f 106
OY 125 LGLSLSLVLDPLSLSYSHFLLEKFGFNKLTVOIMITDMKSLTLAAIAGC---PILX 181
Db 107 VVLSILTLVLCLPLDWIgyvslg-ymslqtaswakdqyisfswisfpitlclvlyf 165
OY 182 LFLKLFDFKFPDFLWYIMW---FLFVVQILAMTIIIPVFMFMFKFPLEDEGLKKSIE 237
Db 166 wlkthekkwlyawllcvrfsflffifig-----pylidplyndfplrkkeleskll 218
OY 238 SLADRVGPEPLDKIEVIDGSRSSHSNAYFTGLPFTSKRIVLPDVLVNSNSTDITAVLAH 297
Db 219 eladeanlpadhvyemsektnalnayvtgig-ankrtivlwtclnklddseillfimg 277
OY 298 EIGHOKNHIYVMVIFESQHTFLIFSL-----FTSIRNTSYNNTFGFLEKSTGSFVDP 352
Db 278 emghyvmkh-----vyiglagyllvslagfyvdiklyktrvlrslsmfhlegrhdaalp 332
OY 353 VITKEFPIITIGFMLENDLTPLECAMQFVMSLSRTHFYQADAVAKKL-GYKQNLGRALI 411
Db 333 I-----llllsvlsfavtfsna-----vsrygenkadqylteltenreaavkftg 379
OY 412 DLQIKNSTMNVDPLYSYSHPTLAERSTALDYVESEKKN 453
Db 380 dlavtqlsqvdpvlyvkllfrgshpsimeriqh-----aekeen 417

RESULT 14
AAW20512 standard; Protein: 192 AA.
XX
AC AAW20512;
XX
DT 14-JUL-1997 (first entry)
XX
DE H. pylori cell envelope protein 4714375.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteriophage; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 51 /label= unknown
FT /note= "encoded by SGC"
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
XX
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR ) ASTRA AB.
XX
PI Berglindh OT, Smith D, Mellgaerd BJ.
XX
DR WPI: 1997-052306/05.
XX
DR N-PSDB: AAT67669.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 56; Page 673; 1481pp; English.
XX
CC This sequence is a H. pylori cell envelope protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,

```

```

CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. No
CC identity likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 192 AA;
XX
Query Match 11.7%; Score 274.5; DB 18; Length 192;
Best Local Similarity 33.2%; Pred. No. 5.2e-21;
Matches 72; Conservative 40; Mismatches 74; Indels 31; Gaps 8;
OY 239 LADRVGPEPLDKIEVIDGSRSSHSNAYFTGLPFTSKRIVLPDVLVNSNSTDITAVLAH 298
Db 1 mmokvgyfsgqilfmdaskrdgrlnayf99ig-knkrrvllfcllsvkgtexlallghe 59
OY 299 IGHWO-KNHIYVMVIFESQHTFLIFSLFTS---IYRNTSFYNTGFFLEKSTGSFVDPV 353
Db 60 lghfkxkdlknlglimggl-lalvfallahlpplvfe-----gfnvsqpsall--- 107
OY 354 ITKEFPIITIGFMLENDLTPLECAMQFVMSLSRTHFYQADAVAKKLGYKQNLGRALID 412
Db 108 -----llllflpvsfyampI-----lgfstrkneyackfgaslskctakalvs 155
OY 413 LQIKNSTMNVDPLYSYSHPTLAERSTALDYVE 449
Db 156 lvenkafpshfylvflhftnpplerlkaldyele 192

RESULT 15
AAV05502 standard; Protein: 211 AA.
XX
AC AAV05502;
XX
DT 05-JUL-1999 (first entry)
XX
DE Bacillus subtilis metalloprotease YHFN polypeptide (aa198-426).
XX
KW Metalloprotease; protease; YHFN; textile; detergent; feedstuff;
KW animal feed; host expression system.
XX
OS Bacillus subtilis.
XX
FH Key Location/Qualifiers
FT Region 80..84 /note= "HEXXH metalloprotease motif"
XX
PN WO9914341-A2.
XX
PD 25-MAR-1999.
XX
PF 08-SEP-1998; 98WO-US18677.
XX
PR 15-SEP-1997; 97GB-0019637.
XX
PA (GENV ) GENENCOR INT BV.
XX
PA (GENV ) GENENCOR INT INC.
XX
PI Estell DA.
XX
DR WPI: 1999-244036/20.
XX
XX
CC New metalloproteases derived from gram-positive microorganisms
CC Claim 15; Fig 3; 54pp; English.

```

xx This polypeptide corresponds to amino acid residues 198-426 of
CC YHFN (see AAY05501), a novel metalloprotease (MP) of the
CC Gram-positive microorganisms *Bacillus subtilis* strain I-168.
CC The MP polypeptide shows homology to *Bacillus stearothermophilus*
CC subtilisin J. The invention relates to the identification of
CC novel MPs in Gram-positive microorganisms, especially *B. subtilis*,
CC *B. stearothermophilus*, *Bacillus licheniformis* and *Bacillus*
CC *amyloliquefaciens*. The invention also provides methods for the
CC production of MPs in host cells as well as for the production of
CC heterologous proteins in a host cell in which all or part of the MP
CC gene is mutated or deleted. Gram positive MPs have applications in
CC the textile industry, in cleaning compositions (claimed), and in
CC animal feed. Claimed cleaning compositions can include the present
CC polypeptide.
xx

Sequence 211 AA:

Query Match 9.6%; Score 225.5; DB 20; Length 211;
Best Local Similarity 28.0%; Pred. No. 1e-15;
Matches 68; Conservative 43; Mismatches 85; Indels 47; Gaps 8;

QY 217 IMPMENKFPLEDEGLKKSIESLADRVGFPPLKIEVIDGSKRSSHSMAYFTGLPTSKRI 276
Db 1 idplyndfyplknkelskleleadeanipadhvevmsektnainayvtgig-ankri 59
QY 277 VLPDPLVNSNSDEITAVLAHEIGHQKNIHNMVIFSQLHTFLFSL-----FTSIYRN 331
Db 60 vldtclnklddsellfmgmhgmhkh-----vylglaagllvslagfvyldklykr 114
QY 332 TSFYNTFGFLEKSTGSEFVDPVITKEFPITIGFMLEFNDLTPLECAMQFVMSLISRTHEY 391
Db 115 tvrltsmfhlgr-----hd--tpfsna-----vstrygen 143
QY 392 QADATAKKL-GYKQNLCLRALIDQIKNLSTMNVDPLYSSYHYSHPTLAERSTALDYVSEK 450
Db 144 kadqygieltenreaavkftfgdlavtqslsqvdpvlykifrgshpsimeriqh---aek 199
QY 451 KKN 453
Db 200 een 202

Search completed: February 12, 2002, 09:27:24
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:13 ; Search time 43.94 Seconds

(without alignments)
1507.996 Million cell updates/sec

Title: US-09-165-460A-2
Perfect score: 2342
Sequence: 1 MEDKTIIDHPNIPWKLIIIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_protist:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1085.5	46.3	462	3	09C2C0
2	819.5	35.0	459	10	09M139
3	730.5	31.2	451	5	09V7W7
4	654	27.9	316	10	004602
5	612.5	26.2	442	5	09XVES
6	501.5	21.4	395	2	09P14
7	459	19.6	407	2	09ZKE0
8	454	19.4	407	2	025144
9	406.5	17.4	447	5	09V7W6
10	378	16.1	428	1	09HS2
11	372.5	15.9	426	5	09V7W5
12	241.5	10.3	504	5	09V7W5
13	197	8.4	338	1	09V0G6
14	152	6.5	291	2	09V0G6
15	148.5	6.3	279	2	09JVI9
16	136	5.8	347	2	09K006
17	135.5	5.8	287	2	09KST9
18	132.5	5.7	760	1	09PA93
19					09UX10

20	129.5	5.5	304	2	09K9B6	09K966 bacillus ha
21	128	5.5	2298	8	09B1K6	09B1K6 lotus japon
22	125	5.3	316	2	09A5E1	09A5E1 caulobacter
23	123	5.3	315	1	029337	029337 archaeoglob
24	122	5.2	3119	5	025857	025857 plasmidium
25	120.5	5.1	289	1	09U2K3	09U2K3 pyrococcus
26	120.5	5.1	298	2	09A1D5	09A1D5 streptococ
27	118	5.0	1436	10	09A1D5	09A1D5 quillardia
28	117.5	5.0	324	2	09C1I2	09C1I2 lactococcus
29	115.5	4.9	485	1	09H1C1	09H1C1 thermoplas
30	112.5	4.8	638	5	002260	002260 caenorhabd
31	110.5	4.7	355	5	017519	017519 caenorhabd
32	110.5	4.7	2280	8	09MER2	09MER2 caenorhab
33	109.5	4.7	669	9	P74499	P74499 synchocyst
34	108.5	4.6	913	5	097272	097272 plasmidium
35	107.5	4.6	236	8	09T241	09T241 phytothor
36	106.5	4.5	521	13	09Y164	09Y164 fundulus he
37	106.5	4.5	645	12	092491	092491 bombyx mori
38	106	4.5	842	10	049423	049423 arabidopsis
39	105.5	4.5	303	2	P96088	P96088 thermococ
40	105.5	4.5	478	8	09ZY24	09ZY24 pedionomas
41	105	4.5	572	2	099X20	099X20 staphylococ
42	105	4.5	1191	2	09V0T8	09V0T8 drosophila
43	104.5	4.5	350	2	030362	030362 thermococ
44	104.5	4.5	499	2	09C49	09C49 lactococcus
45	104	4.4	333	11	09JKF0	09JKF0 rattus norv

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	462 AA.
09C2C0	09C2C0			
AC	09C2C0			
DT	01-JUN-2001 (TREMREL, 17, Created)			
DT	01-JUN-2001 (TREMREL, 17, Last sequence update)			
DT	01-JUN-2001 (TREMREL, 17, Last annotation update)			
DE	PROBABLE ZINC METALLO-PROTEASE.			
GN	B1N2.80.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes.			
OC	Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_Taxid=5141;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Schulte U., Algen V., Hohnselt J., Brandt P., Farlmann B., Holland R.,			
RA	Nyakatura G., Mewes H.W., Mannhaupt G.;			
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	German Neurospora genome project.			
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AL513444; CAC28689.1;			
KW	Protease.			
SO	SEQUENCE 462 AA; 52825 MW; 6B4ED30AC028EC CRC64;			
Query Match	46.3%; Score 1085.5; DB 3; Length 462;			
Best local similarity	47.3%; Pred. No. 3.2e-69;			
Matches	212; Conservative 79; Mismatches 140; Indels 17; Gaps 3;			
Db	4	LKTIDHPNIPWKLIIISFSTIAQSFESYLYTRQOKSERKLPVLEDEIDDEFHNSR	63	
Db	7	LARFLDRPLFPWKLLIFESFLANFYIEGFLGRQVAKTKRPVLEHEVSQEVFDSQ	66	
Db	64	NYSRAKAKFSIFGDVYNLAOKLFTKYDLPFRIMHMAVSLNAVLPVFHMAVSTAQSLC	123	
Db	67	AYGRKAKFEFFSGIYGIONILFFQPDVLRKMSFAGNLVRFAPARFS--GELSGSIV	124	
Db	124	FLGLISLSTLVDPLPLSYSHFVLEKRGFNKLVQVMTIMKISLTALVAGGPIYLF	183	
Db	125	FVLFSFVVISQILSLPTSLYHTEFVLEKRGFNKSAKLTWTKIKISLFLVLTPIIAGF	184	

QY	39	QKLSSETLPLVLEDEIDDEFTHKRNRSRAKAKSFIFGDVYNNIAOKLVFETKYDLFPKRIWM	98
Db	35	OKLCE--KPYL--LPQKDYEAAGNVAIRKMOJLSTIIQLD--GVIFAGW-VFGLTH	84
QY	99	M-AVSLIANAVLPFRFMWSTVAOSLCFGLGLSLSLVDPLPLSYSHVLEKEFGENKL	156
Db	85	LEDLTHYLN-LP--ETLGLVAFALLFLAIQSVLA---LPISYTTTMDLKEFGESKV	135
QY	157	TVQLMTIDMIKSLTLAAIGPRLYLFLKTFDEKFPPTDFLWYIWF--LFWQJLMTIIP	214
Db	136	SLSLFPDFDEFGLLLTGLGVGLLTLYTLIMIIENHV--WEISFVYVFFMILLANFYP	192
QY	215	VFIPIWENKFTPELDEGLKKSIESLADVGFPDLKIFVIDGSKRSSHSNAYFTGLPFTSK	274
Db	193	K-IQLTFNOFPPLNRPDLLESQIESMDKVGFKSGCIFPMADSKRDGLNAYFEGGLG-KNK	250
QY	275	RIVAFEDPLVNSNSDELTAVALHIGIMO-KNHVYVNIIFSOLHTFLFESLFTS---LY	329
Db	251	RVVLEFDLTLVAVGTGGLLALIGHLGHFKNKDLKNLGLINGCL-LALVFALLAHLPPLVF	309
QY	330	RNTSFYVTGTFEFLKSTGSEFVDVPTTKEPFIIGFM-LFNDLTPLECAMQFMSLSIRT	388
Db	310	E-----GNVSGQTFASLI-----TILLFLPVEFVFAMPL-----IGFPSRK	346
QY	389	HEYQADYAKKLGKYNKLCRALIDLQIKNLSITMNVDPLEYSYHSHPTLERSTALDYVS	448
Db	347	NEYADKFGASLSSEKETLAKRLAVSIVENKNAFPYSHPFVYVLFHTHPPLERUKADYEI	406
QY	449	E 449	
Db	407	E 407	
RESULT	8		
025144			
ID	025144	PRELIMINARY;	PRT; 407 AA.
AC	025144		
DT	01-JAN-1998	(TREMBLrel. 05, Created)	
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	ZINC-METALLO PROTEASE (YU8117M).		
GN	HP0382.		
OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;		
CC	Helicobacter.		
OX	NCBI_TaxID=210;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=26695 / ATCC 700392;		
RX	MEDLINE=97394467; PubMed=9252185;		
RA	Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,		
RA	Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,		
RA	Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,		
RA	Loftus K., Richardson D., Dodson R., Khalak H.G., Glodek A.,		
RA	McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,		
RA	Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,		
RA	Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Matthey L., Wallin E.		
RA	Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,		
RT	Venter J.C.		
KT	"The complete genome sequence of the gastric pathogen Helicobacter		
RL	pylori."		
DR	Nature 368:539-547(1997).		
DR	EMBL; AE000555; AAD07451.1;		
DR	MEROPS; M48. 008; -		
DR	TIGR; HP0382; -		
DR	InterPro; IPR001915; Peptidase_M48.		
DR	Pfam; PF01435; Peptidase_M48; 1.		
FW	Hypothetical protein; Protease; Complete proteome.		

```

QY 39 OKSEETKAPRLDEIDEIDEPHKSRRNSPAKAKSIJCGDVYNLAOKUVEFKYLDAPKIMH 98
Db 35 OKLE--KPL--LQOKDIEENGAINRMOJSLISQILD--GIIFGW--VEFGLTH 84
QY 99 M--AVSLINAVLPVRFHMVSTVAOASLCPGLGSLSLSTYDLPULPSYSHFVLEEKFGENKL 156
Db 85 LEDLTHYN--Lr---ETLGIVFALLFLIAQSLA---LPISYTYMHLDDPEFGSKY 135
QY 157 TVQJLNTDMIKSLTLAVAIIGRPILYFLKLTDFKFPDLEWYIMV--LEVVOJLAMTIIP 214
Db 136 SLSEFFKDFEFGSLSTLSVGLLTYLTIMIEIHEH--WEISFEVVEFMTLANLEYP 192
QY 215 VFIIMPENKFPLEDGELKKSIESLARBVGPRDKIPVIGDSRRSSNMYFGLDPTSK 274
Db 193 K-IQGLEQEPFLNNRDLSEIIEBMMKRVGKSGEIPVAMSKRDGLNLYFGGLG-KNK 250
QY 275 RIVLEFDLPVNSNSDEITAVLAHEIGHMOKNIYINWYIFQSLHTFLFSLFSTS---IYR 330
Db 251 RVLEFDPLIKSVGTEGLLALIGHGLGHPKKNDLKSIGINGGLLAVFALIALHPLRYVE 310
QY 331 NTFSEYNTGFPLEKSTGSEFVDVPVITKEFPIIGFM-LFNLDLPLPCEAMQFVNSLSRTH 389
Db 311 -----GFNVQSPASLI-----ALLLEPLPVESEYAMPL-----IGFESRKN 347
QY 390 EYQADAVAAKRGYQONLCRALIDQJINLSIMWNDPLXSYHNSHPILARSTRALDYSE 449
Db 348 EYNDKRCGASLSSEKVEYLAALAVSYVSNKAPVYSHPRVYLFHPTPLERLAKLADYE 407

```

RESULT	9			
09V7W6		PRELIMINARY:	PRT:	447 AA.
AC	09V7W6			
DT	01-MAY-2000	(TREMBlrel. 13, Created)		
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)		
DT	01-JUN-2001	(TREMBlrel. 17, Last annotation update)		
DE	CG9001	PROTEIN.		
GN	CG9001.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-BERKELEY.			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Abn K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,			
RA	Wahl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,			

RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Rubin G.M., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003804; AAF57924.1;
 DR FlyBase: FBgn0034174; CG9002.
 DR InterPro: IPR001915; Peptidase_M48.
 DR Pfam: PF01435; Peptidase_M48; 1.
 SQ SEQUENCE 426 AA; 49971 MW; 0760ECF84F291927 CRC64;

Query Match 15.9%; Score 372.5; DB 5; Length 426;
 Best Local Similarity 25.1%; Pred. No. 9.8e-19;
 Matches 110; Conservative 81; Mismatches 185; Indels 63; Gaps 13;

QY 33 LTVRQYOK-LSEFTKLPVLEDEIDDETHKSNYSRAKAKPSIPGDVYNLAKVFIKXD 91
 DB 3 LTRKQALVCLNAINVPEERGIIPETIHRARIELHTELDIMWYLLDLITLLELLIG 62
 QY 92 LEPKIMHNAVSLNANVLEVRFMHSTVAQSLGILSLSTLVLDPLSYSHFLEEKF 151
 DB 63 FYFPLMSLSATLQKITSGEIMI-----TLIFVYIITVICIRPLVILYKCLELRY 116
 QY 152 GFN-KLTVQMLTMDIKSLTAYAGPIL-----YLPKIDKFPDPLWYIM 199
 DB 117 GMSGFPPWYLCIGAMISLLSQLVLPALAAIVSVKFIGYF-----FLWF-W 165
 QY 200 VFLFVOLLAMTIPVFLIMPENKFTPLEDG-ELKKSIESLADRVGFPLDKIFVLDGSR 258
 DB 166 LFWATFTLLVFLFDPYCCIPCIIGROVVLPEGALTMEYKRVCDVVGFPKRVFIIR-TYT 224
 QY 259 SSHSNAYFTGLPFTSKRIIVLEDTLVNSSTD-----ETVLAIEIGHM 302
 DB 225 MQYSNAYRYG-SCCLAKRIIVFTLLNKGKEPNEIHPYEVGGLINIOVAGVCHGLGM 283
 QY 303 OKNHTIVNVIISQHTPLFLISLTISYRNTSYFNFGFFLEKSTGSEVDPVITKEPPII 362
 DB 284 KHGHFYKATITMKIHFFITMGLGFFPHSPOLYMAVGF-----EPGV---MPLIV 330

QY 363 GFML-FNDLTPLECAMQFVMSLSRTHGYADAVAKKIGYKONICRALID:QINKLSTM 421
 DB 331 GFTIYLKALRPYUPLTANVLMNLNLRREYADKFAHRNGYSIOLRMALVKIYADHMSPP 390
 QY 422 NNDPLYSSYHVSHTPLAER 440
 DB 391 VYDQCYARHNHTHTPIIDLR 409
 RESULT 12
 ID 09VTH6 PRELIMINARY; PRT: 504 AA.
 AC 09VTH6:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG7573 PROTEIN.
 GN CG7573.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Rubin G.M., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003804; AAF50072.1;
 DR FlyBase: FBgn0034153; CG7573.
 DR InterPro: IPR001915; Peptidase_M48.
 DR Pfam: PF01435; Peptidase_M48; 1.
 SQ SEQUENCE 504 AA; 57737 MW; 892022B675E6D49E CRC64;

Query Match 10.3%; Score 241.5; DB 5; Length 504;
 Best Local Similarity 20.7%; Pred. No. 2.3e-09;
 Matches 93; Conservative 85; Mismatches 190; Indels 81; Gaps 17;

28 SPESLITROYOKLSELP-VEDELDDETFHKSRYNRSAKAFSIFGDVYNAQKLV 86
 85 SFHLICROKLCKRTSVPRKMEGIIISOEFAQSKDKOHLASLEIF---NIAIDTL 140
 QY 87 FIKYDLF---PKIWMHNVSLNANLVPVRFHNVSTVAASLOCFGLSLSLTVLDPLSY 142
 DB 141 YSCDLVLYCTLAFLMKLVGM-----YHASTWLNANVFMVFTSYLVKRLPSLEY 192
 QY 143 SHFVLEEKFGFN-KLTVOLMTDMIKSLTAVAGPILYFLKIFDKPPTDFL--WYIM 199
 DB 193 EKLVDPRYVNDPEKTPPL--LGILCALVFAV--IPLAIFMAI-----HMLESEWFT 243
 QY 200 VFLEP---VOLANTTIPVTFIMPMFNKFTPEDEGLKSLIESLADRVGFLDKTFEIDG 255
 DB 244 LEVWLGVLGLSVLAVLAFGLFGVPLGKSRKMNSTDMNSLKAVALDENFP--GRVYV-- 300
 QY 256 SKRSHSNAYFTGLPFT-----SKRIVLFDTL-----VNSNSTD-----EIT 292
 DB 301 -----HFFHVGRTAVNMGCCCLRIDHDKMNRGSSDDPDMGOMAGLDEOLA 353
 QY 293 AVLAHEIGHQKNIHVNIVFSQHTPLIFSLFTSYRNTSFYNTFGFLEKSTGSFVDP 352
 DB 354 AFVAHQHLMQDMHAKLALIFYLILYLLFGICNRMWILYAAAG----- 401
 QY 353 VITKEPFIITIGFMDLTLPL-ECAMQFVMSLSRTHEYQADAVAKKLGKONICRALI 411
 DB 402 --TTFYPSVGFWLVKYLMPYIHDISTVIFECIRHFEVADAVNRGGLPRAALL 459
 QY 412 DLQIKNSTMNVDPLSYSHYSHTPLAER 440
 DB 460 KLPSDDEFFPYVDOCYLMMHRLRPSVLOR 488

Query Match 8.4%; Score 197; DB 1; Length 338;
 Best Local Similarity 22.3%; Pred. No. 2e-06;
 Matches 86; Conservative 66; Mismatches 133; Indels 108; Gaps 18;

96 IMHNAVSLNANLVPVRF--HMVSTVA-----OSLCFLGLSLSLTVLDPLSY 142
 DB 2 LMEHLI-LFGMILPVLPMHRIKELRGKRGKRMKFEKVLFEALIIISLVFSLAGIIGF 60
 QY 143 SHFVLEEKFGNKLTVOL--WITDMIKSLTAVAGPIL--VLFKIFDKPPTDFLWY 197

DB 61 FDFAMSFIEPLDKIIVLPVLPVSTLLITRELGEKIRRGDYIVLAVF----- 110
 QY 198 IMVFELVY---QLAMTIIIVTFIMPMFNK-----TPEDEGLKSTISL 239
 DB 111 -IVFEFLVLSFSIIPMYVPVLYLMLSPAELESRIYRRFFHGTPM--SEGLRAKABEL 168
 QY 240 ADVRFEPDKIFVIDGSRSSHSNAYFGCLPFTSKRIYLPFTLVNSSTDEITAVLAHEI 299
 DB 169 CRANVANVEEYIID---EERIGAFVYGM--KGTITITGALIEKLNEMELLAVIAHEL 222
 QY 300 GHMOKNHIHVNIVFSQHTPLIFSLFTSYRNTSFYNTFGFLEKSTGSFVDPVITKEFP 359
 DB 223 GHVKRRHALKREL-----ALVFGL-----SLP 244
 QY 360 IIGFMNLNDLTLPECAMQFVMSL-----ISRTHEYQADAVAKKLGKONICRAL 410
 DB 245 -TIGYMGDI--PL--FLSFVMSIALVFYSTFOIAKKEIDADRFAASLVGPIVIVAL 299
 QY 411 IDLQIKNSTMNVDPLSYSHYSHTPLAERSTALD 445
 DB 300 EKVYKKEGLPRTPRWYNIH--SHPSLEERVRALE 333

RESULT 14
 ID 091013 PRELIMINARY; PRT: 291 AA.
 AC 091013;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE HEAT SHOCK PROTEIN HTPX.
 GN HTPX OR PA2830.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAOI.
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larrick K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
 opportunistic pathogen."
 RL Nature 406:959-964(2000).
 DR EMBL: AE004709; AAC66218.1;
 DR InterPro: IPR001915; Peptidase_M48.
 DR Pfam: PF01435; Peptidase_M48; 1.
 KW Complete proteome.
 SQ SEQUENCE 291 AA; 31593 MW; EC2B68695E5EFAAB CRC64;

Query Match 6.5%; Score 152; DB 2; Length 291;
 Best Local Similarity 24.4%; Pred. No. 0.0026;
 Matches 72; Conservative 46; Mismatches 123; Indels 54; Gaps 11;

QY 179 ILVYFLKIF--DKPPTFLWYIMVFLVYQILAMTIIIVTFIMPMFNKFT-----PLE 228
 DB 17 IASTIKLGVDRFTGQYVGLVFCVAVG--FAGSLVSLFSKMAKASTGTETVISOPT 75
 QY 229 DGE--LKKSIESLADRVGFLDKIFVIDGSRSSHSNAYFTGLPFTSKRIYLPFTLVNSN 286
 DB 76 RHEQMLQTYEELSRGAKIMPEV---GIPAYEANAFAFGMKNKNDALVAVSOGLIERF 131
 QY 287 STDTEITVLAHEIGHQKNIHVNIVFSQHTPLIFSLFTSYRNTSFYNTFGFLEKST 346
 DB 132 SPDEKAVLAHEIGHVANGDVLTALIO-----GVVNTVFVFARIF 173
 QY 347 GSFVDPVITK--EPPITIGFM--LFNDLTLPECAMQFVMSLSRTHEYQADAVAKKLG 402

Db 174 GNFYDKALKNKEDGPGIGYFATIFAEV--LGLIASTIVWFSRRREFRADAGAHLAG 231
Y 403 KQNCRALIDQ-----IKNLSTMNVDPLYSSTHY-----SHPTLAERSTAL 444
Db 232 TGAIIAALQRLRSQGVVPMPTLMAFGINGLKGLAGLMSHPLEDRIEAL 286

RESULT 15

09JVI9
ID 09JVI9 PRELIMINARY; PRT: 279 AA.
AC 09JVI9;
DT 01-OCT-2000 (Tremblrel, 15, Created)
DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
DE PUTATIVE MEMBRANE-BOUND ZINC METALLOPEPTIDASE (EC 3.4.24.).
GN HTPX OR NMA1031.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491".
RT Nature 404:502-506(2000).
RL EMBL: AL162754; CAB84300.1;
DR InterPro: IPR001915; Peptidase_M48.
DR Pfam: PF01435; Peptidase_M48.1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 279 AA; 30174 MW; 713442CD6CAC015E CRC64;

Query Match 6.3%; Score 148.5; DB 2; Length 279;
Best Local Similarity 25.7%; Pred. No. 0.0044;
Matches 59; Conservative 42; Mismatches 88; Indels 41; Gaps 10;

QY 228 EDGLKKSIESLADRVGFPDLKIFVIDGSKRSSHSHNAYFTGLPFTSKRIVLFDLTVNSNS 287
Db 75 EEAMLMTVEAARQWNLKTPVAIY---HSPENAFATGASRNSSLIAVSTGLDHT 130
QY 288 TDEITAVIAHEIGHMOKNHIYNM-VIFSQLTFLIFSLSIYRNSTSYNTFGFLEKST 346
Db 131 RDEVEAVIAHEMAHYNGDMVTLTLLIOGVNTFYVF--LSRIIANLIRNNDG---SOSQ 185
QY 347 GSFVDPVITKEPFIIGFMLENDLLTPECAMQFVMSLSIRTHGYOADAVAKKIGYKQNL 406
Db 186 GTTF--LVSMYFQILFGFL-----ASLIVMFSRQREYRADAGAKLVGAPKM 231
QY 407 CRAL-----IDLQKNLSTMNV----DPLVSSYHSHPTLAERSTAL 444
Db 232 ISALQRLKGNFVDLP--EEMNMGIAGTDRDLSL---THPSLDNRIRAL 276

Search completed: February 12, 2002, 09:32:01
Job time: 348 sec

THIS PAGE BLANK (1.3PTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:25:34 ; Search time 21.66 Seconds
(without alignments)
470.637 Million cell updates/sec

Title: US-09-165-460a-2
Perfect score: 2342
Sequence: 1 MEDLKTILDPNIPMKLIIS.....HPTLAEKSTALDYVSEKKKN 453

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2-6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2-6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2-6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2-6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2-6/ptodata/2/1aa/PTUS-COMB.pep.*
6: /cgn2-6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	653	27.9	474	3	US-09-022-699-2
2	105	4.5	3200	2	US-08-477-451-8
3	100.5	4.3	1226	2	US-08-540-804-12
4	100.5	4.3	1226	2	US-08-218-265-12
5	100.5	4.3	1226	3	US-08-521-872-12
6	100.5	4.3	1226	4	US-08-590-399-12
7	94	4.0	1876	2	US-08-619-554-2
8	93.5	4.0	485	2	US-08-477-451-22
9	92.5	3.9	284	1	US-08-118-270-67
10	92.5	3.9	284	5	PCT-US93-08528-67
11	92.5	3.9	1030	4	US-09-091-117-2
12	92	3.9	1285	1	US-07-382-945-2
13	92	3.9	1285	2	US-08-453-141-2
14	92	3.9	1285	3	US-08-293-314-2
15	90	3.8	485	1	US-07-991-8678-42
16	90	3.8	485	2	US-08-544-332-42
17	89	3.8	194	4	US-08-959-004-9
18	88.5	3.8	498	1	US-08-457-274A-24
19	88.5	3.8	498	5	PCT-US95-05758-24
20	88.5	3.8	1786	2	US-08-477-451-16
21	88	3.8	1045	1	US-07-596-467-6
22	88	3.8	1045	1	US-07-934-374-6
23	88	3.8	1045	1	US-07-783-861C-6
24	87.5	3.7	1781	2	US-08-477-451-11
25	87	3.7	440	1	US-08-307-499-15
26	87	3.7	440	4	US-09-299-268-15
27	87	3.7	599	1	US-08-295-814E-11

28	87	3.7	599	1	US-08-240-783B-4	Sequence 4, Appl1
29	87	3.7	599	3	US-09-084-813-4	Sequence 4, Appl1
30	87	3.7	599	4	US-09-343-361-11	Sequence 11, Appl1
31	87	3.7	599	5	PCT-US92-09662-4	Sequence 4, Appl1
32	86.5	3.7	502	1	US-08-484-840-3	Sequence 3, Appl1
33	86.5	3.7	502	1	US-08-483-094-3	Sequence 3, Appl1
34	85	3.6	599	1	US-08-301-722A-5	Sequence 5, Appl1
35	84	3.6	432	4	US-09-370-253-12	Sequence 12, Appl1
36	84	3.6	1394	4	US-09-213-053-2	Sequence 2, Appl1
37	83.5	3.6	277	1	US-08-118-270-68	Sequence 68, Appl1
38	83.5	3.6	277	5	PCT-US93-08528-68	Sequence 2, Appl1
39	83.5	3.6	1956	4	US-08-843-417-2	Sequence 2, Appl1
40	83	3.5	1079	3	US-09-136-652-2	Sequence 4, Appl1
41	83	3.5	1895	2	US-08-619-554-4	Sequence 2, Appl1
42	82.5	3.5	355	2	US-08-458-970A-2	Sequence 2, Appl1
43	82.5	3.5	1079	1	US-08-455-343A-55	Sequence 55, Appl1
44	82.5	3.5	1079	2	US-08-223-305C-55	Sequence 2, Appl1
45	82	3.5	620	1	US-08-301-722A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-022-699-2
Sequence 2, Application US/09022699
Patent No. 6060277
GENERAL INFORMATION:
APPLICANT: KIKLY, KRISTINE
APPLICANT: SOUTHAN, CHRISTOPHER
APPLICANT: KNAB, ANNE
TITLE OF INVENTION: Human AFCl
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.699
FILING DATE: 12-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 97304440.7
FILING DATE: 12-JUN-97
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70380
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-022-699-2
Query Match 27.9%; Score 653; DB 3; Length 474;
Best Local Similarity 34.7%; Pred. No. 1,1e-59;
Matches 161; Conservative 73; Mismatches 166; Indels 62; Gaps 8;

```

QY 22 FSIQSFESLYLTYRQ---YQKLSSTKLPVLEDEIDETPHKSRNYSRAKAKFSIFGDV 78
DB 26 FSWTYLWMTFLAQRORRYK--TTHVPELGGQIMDSSTFHSKRIYQDKSTFWSGL 83
QY 79 YNLQKLVYIKIDLPKTHMAVSLNNAVLPVPHMVSVVAVOSLCLGLSSLSLTVLDP 138
DB 84 YSEETGDTLLFLGGIPYLMRLKRGRCGY--AGFGPEYITQSLVFLLATLFSALTGLP 140
QY 139 LSYSHFVLEEKGFNKLVLQIMDMIKSLFLAVAGSPILYLFKIDPKPTDFLWY 198
DB 141 WSLYNTFVEEKHGNQQLGFFMDAKKVVYTOCILLPVSSLLYTIKIGDYFTYA 200
QY 199 WVELEVOVLAMTIIPEVPMNFKPTLEDELKKSIESLADRYGFPDLKIFVIDGSR 258
DB 201 WLFYLVSLVLTIVADYIAPLEFKETPLPECKLKEIEVMAKSIDPILTKYVVEGSR 260
QY 255 SSHSNAYFTGLPFTSKRIYLFDTLVNSN----- 287
DB 261 SSHSNAYFTGF-FKNKRIVLPDTLLEYSVLNKDIOEDSGMEPRNEEGNSEIRAKYKN 319
QY 288 -----TDEITAVLAHEIGHMOKNHVNVIFSOHLTFLEISFTSYNTSFYNTFGF 341
DB 320 KQGGCKNEEVLAVLGHGSHW-KMDIOSKISLAREFPFVFLAVLGRKELFAFGRY 378
QY 342 LEKSTGSPVDVPTKEFFPIIGFM-LFNDLTLPLCAMQFVMSLSIRTHEYOADAVAKKL 400
DB 379 DSQ-----PILIGLLIIFQIFSPYNEVLSFCLTVLSRRREFQADAFAKKL 424
QY 401 GYKQMLCRALDLOIKNLTSMNVDPVLYSSYHSHPTLAERSTAL 444
DB 425 GKAKDLYSALIKLNKDNLGFPVSDWLFSMWHYSHPPLLERLOAL 468

```

RESULT 2

US-08-477-451-8

```

: Sequence 8, Application US/08477451
: Patent No. 5928865

```

```

: GENERAL INFORMATION:

```

```

: APPLICANT: Covacci, Antonello
: TITLE OF INVENTION: Helicobacter Pylori CagI Region
: NUMBER OF SEQUENCES: 46

```

```

: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA

```

```

: COUNTRY: USA

```

```

: ZIP: 94608-2916

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,451

```

```

: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.

```

```

: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0335, 002

```

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 510-601-2708

```

```

: TELEFAX: 510-655-3542

```

```

: INFORMATION FOR SEQ ID NO: 8:

```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 3200 amino acids

```

```

: TYPE: amino acid
: STRANDEDNESS: single

```

```

: TOPOLOGY: linear
: MOLECULE TYPE: protein

```

US-08-477-451-8

```

Query Match 4.5% Score 105; DB 2; Length 3200;
Best Local Similarity .24, 8%; Pred. No. 0.25;
Matches 102; Conservative 55; Mismatches 152; Indels 102; Gaps 23;

QY 38 YQKLSSTKLPVLEDEIDETPHKSRNYSRAKAKFSI-----FGDY-----N 80
DB 1274 FRNKNTIIPGYSILRVLEIFNLSEFFLFLAQVYSIILCKFHEIYFVLDLFLICS 1333
QY 81 LAQKLVF--IK-----YDLFKIMHVAVS-----LNAVLPV----- 110
DB 1334 LACELVFEVEIKTRVNVPELFTCYDA-TAVFHMHNCSHIVALLVFTINGVPIPSGR 1392
QY 111 -----RHVNSYV---AQSLCLGLSSLSLTVLDPVLSYSHFVLEEKGFNKL 156
DB 1393 TTIAPIQNEFFRYALIMPVHRRQSV-IVNQLIELVIGINVDSLNHHT- AKSEPKI 1450
QY 157 TVQLMTDMIKSLFLAVAGSPILYLFKIDPKPTDFLWYIMVFLVQVLAMT--IP 214
DB 1451 TPFI-----KISFRALYVGGSGYIMDMGMWFLIDCFDIR-RLFFVYKYEKNIQPIP 1505
QY 215 VFIPEMFKPTLEDELKKSIESL-ADRYG-----PLDKIFVID--GSKR 258
DB 1506 TSRLSIFFCFLPFQGLIKHAHELLIRSDLIPIYNDCKRLEFGQKLRFLTIKRGLYD 1565
QY 259 SSHSNAYFTGLPFTSKRIYLFDTLVNSNSTDEITAVLAHEIGHMOKNHVNVIFSOHLT 318
DB 1566 IFHKISDSNGVDF--RIVLMGHPCENSVTLIPFLKALGCRYINEIVFDIITALIDK 1623
QY 319 FLIFLSFTSYNTSFYNTSFYNTFGFLEKSTGSPVDVPTKEFFPIIGFMLE 369
DB 1624 VKILLFGRIF--DRFIGVF-FPLRDSFKSL--PIV-----LIND 1661

```

RESULT 3

```

: US-08-540-804-12
: Sequence 12, Application US/08540804
: Patent No. 5919666

```

```

: GENERAL INFORMATION:

```

```

: APPLICANT: Young, Richard A.
: APPLICANT: Koleske, Anthony J.

```

```

: APPLICANT: Thompson, Craig M.
: APPLICANT: Chao, David M.

```

```

: TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
: TITLE OF INVENTION: Transcription and Methods of Use Therefor

```

```

: NUMBER OF SEQUENCES: 39

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millitia Drive
: CITY: Lexington

```

```

: STATE: Massachusetts

```

```

: COUNTRY: USA

```

```

: ZIP: 02173

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/540,804

```

```

: FILING DATE: 11-OCT-1995
: CLASSIFICATION: 424

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/521,872

```

```

: FILING DATE: 21-AUG-1995
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 08/218,265
: FILING DATE: 25-MAR-1994

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia

```

```

: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: WHI94-03A2

```

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1226 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-540-804-12

Query Match 4.3%; Score 100.5; DB 2; Length 1226;
Best Local Similarity 19.5%; Pred. No. 0.18;

Matches 96; Conservative 86; Mismatches 206; Indels 105; Gaps 21;

```
QY 7 ILDHNIWPKLLISGFSIAQSFESLYRYQYKLSERKLPVLEDEIDDEFHKSRYNS 66
DB 374 IIRVPTYIRKLLISGLYLQDSNDKFV--HVQLLINLKISPLMSQYN---MVLRYNM 426
QY 67 RAKAFSIFGDVYNLAOKLVFIKYDLFPKIMHAYS---LNAVLPVRHMYSTVAQSLC 123
DB 427 EYDVAFY---EIFNFDQ-LVEITEOIKRILSNDITNQLSTPLSIKIMVAEWYLSHC 482
QY 124 FLGLISLSTLYDL-----PLSYSHFVLEERGFENKLTVOAMITDMIKSLTAVAI 175
DB 483 -SGILSSVNRVYLLKIFKICIDLEVFNHF-----FKNIEFIVYHQLSDIESLEALM-- 534
QY 176 GGPILYLKLFDPKPTDFLWIMVFLVY-QILAMTIIPV---IMPEN-----KF 224
DB 535 -DILLCYOKLESQFINHILFTKTFIFLYKKVLEKEDVPANVTSFPMFEMKFNKPNPF 592
QY 225 TPLEDGELKKSIESLADRVGFPDVKI-----FVIDSKRSSHNSNAFTGLPPTS 273
DB 593 VLKVNDRIRIELQSVYNDKELKTEKLNKDKSEVLKYVSMINNSNOAVGQTNPFPEVOYN 652
QY 274 KRIVLFDI-LVNSNST-----DEITAVLAHEIGHQWKNHIVNVIQSOLHTF----- 319
DB 653 IREFLHNSIITDNTSKQOKARRNVMLLIATNL----KEYIKFMSIFLKRKDFNKNLI 708
QY 320 -----LIFSLETSIYRNTSFYNTGFFLEKSTGTF-----VDPVITKEFPPIIGFM 365
DB 709 QLISKLTLFEVTONVLEVIIRLLPINLENNDSYGLFLKYHKEQFIKSNEKI---- 764
QY 366 LFNDLLTPL-----ECAMQFVMSLSIRTHEQADAVAKKLYKQNLCAALLDLOIK 416
DB 765 -----LITCYELEKKYHGNDCEINYELIKLIIITYGSS--PKLLATSTKIIIMLLDSVE 818
QY 417 NLSTNMVDPLEYSS 429
DB 819 NSSNLEIDILYYS 831
```

RESULT 4

US-08-218-265-12

Sequence 12, Application US/08218265

Patent No. 5922585

GENERAL INFORMATION:

APPLICANT: Young, Richard A.

APPLICANT: Koleske, Anthony J.

APPLICANT: Thompson, Craig M.

TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/218, 265

FILING DATE: 25-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI94-03

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1226 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-218-265-12

Query Match 4.3%; Score 100.5; DB 2; Length 1226;
Best Local Similarity 19.5%; Pred. No. 0.18;

Matches 96; Conservative 86; Mismatches 206; Indels 105; Gaps 21;

```
QY 7 ILDHNIWPKLLISGFSIAQSFESLYRYQYKLSERKLPVLEDEIDDEFHKSRYNS 66
DB 374 IIRVPTYIRKLLISGLYLQDSNDKFV--HVQLLINLKISPLMSQYN---MVLRYNM 426
QY 67 RAKAFSIFGDVYNLAOKLVFIKYDLFPKIMHAYS---LNAVLPVRHMYSTVAQSLC 123
DB 427 EYDVAFY---EIFNFDQ-LVEITEOIKRILSNDITNQLSTPLSIKIMVAEWYLSHC 482
QY 124 FLGLISLSTLYDL-----PLSYSHFVLEERGFENKLTVOAMITDMIKSLTAVAI 175
DB 483 -SGILSSVNRVYLLKIFKICIDLEVFNHF-----FKNIEFIVYHQLSDIESLEALM-- 534
QY 176 GGPILYLKLFDPKPTDFLWIMVFLVY-QILAMTIIPV---IMPEN-----KF 224
DB 535 -DILLCYOKLESQFINHILFTKTFIFLYKKVLEKEDVPANVTSFPMFEMKFNKPNPF 592
QY 225 TPLEDGELKKSIESLADRVGFPDVKI-----FVIDSKRSSHNSNAFTGLPPTS 273
DB 593 VLKVNDRIRIELQSVYNDKELKTEKLNKDKSEVLKYVSMINNSNOAVGQTNPFPEVOYN 652
QY 274 KRIVLFDI-LVNSNST-----DEITAVLAHEIGHQWKNHIVNVIQSOLHTF----- 319
DB 653 IREFLHNSIITDNTSKQOKARRNVMLLIATNL----KEYIKFMSIFLKRKDFNKNLI 708
QY 320 -----LIFSLETSIYRNTSFYNTGFFLEKSTGTF-----VDPVITKEFPPIIGFM 365
DB 709 QLISKLTLFEVTONVLEVIIRLLPINLENNDSYGLFLKYHKEQFIKSNEKI---- 764
QY 366 LFNDLLTPL-----ECAMQFVMSLSIRTHEQADAVAKKLYKQNLCAALLDLOIK 416
DB 765 -----LITCYELEKKYHGNDCEINYELIKLIIITYGSS--PKLLATSTKIIIMLLDSVE 818
QY 417 NLSTNMVDPLEYSS 429
DB 819 NSSNLEIDILYYS 831
```

RESULT 5

US-08-521-872-12

Sequence 12, Application US/08521872

Patent No. 6015682

GENERAL INFORMATION:

APPLICANT: Young, Richard A.

APPLICANT: Koleske, Anthony J.

APPLICANT: Thompson, Craig M.

APPLICANT: Chao, David M.


```

Db 483 -SGLISSVNRVTLKIFKIFCIDEVEPHNF-----FMWIEPIYVHQLSDIESLEALM-- 534
Qy 176 GCPILYLFKLFKDFPPDFLWYMFLEVY-QVLAATTIIPYF-----IMPMFN-----KF 224
Db 535 --DILLCQKLFSEFINDHILFTFTFIFRYKVLKEDVPAVYNTSMPWPKFPMKNPF 592
Qy 225 TPLEDGELKKSIESIADRVGFFLKI-----FVIDGSRSSHNAVFLGLPFTS 273
Db 593 VLKVDNDRIELQSYNDKELTKERLKNDSVLYKVMSTINNSQAVGQTMNPEVFOVN 652
Qy 274 KRVLFPD-LVNSNST-----DEITAVLAHEIGHQKNHIVMVFSLHFF----- 319
Db 653 IRPLHNSIEIDTMSQFOKARNNVMLLATNL-----KEYIKMSIFLKRKDTNNLI 708
Qy 320 -----LIFSLFTSIYRNISFYVTFEGFLEKSTGSGF-----VDPVITKEPPIIGFM 365
Db 709 QLIISLKLTFTEVTONVIGLEFIIRLPLDINENNDSYGLFLKYHKEQFIKSNFEKI---- 764
Qy 366 LFNOLLPL-----ECAMQVMSLISSTHEYQADAAKKGIVQNLCRALIDQIK 416
Db 765 ----LLTCYELEKKYHGNECINYEILKILITYGSS--PKLLATSTKIMLLANDSVE 818
Qy 417 NLSTMNVDPLXS 429
Db 819 NSSNLEDILYXS 831

RESULT 7
US-08-619-554-2
: Sequence 2, Application US/08619554
: Patent No. 5821353
: GENERAL INFORMATION:
: APPLICANT: DOUGLAS, Cameron M.
: APPLICANT: CHREBET, Gary L.
: APPLICANT: CLEMAS, Joseph
: APPLICANT: EL-SHERBEINI, Mohammed
: APPLICANT: FOOR, Forrest
: APPLICANT: KAHN, Jennifer,
: APPLICANT: KELLY, Rosemarie, - PARENT, S.A.
: APPLICANT: MARRINAN, Jean, - RAMADAN, N.M.
: APPLICANT: MORIN, Nancy, - REGISTER, E.A.
: APPLICANT: ONISHI, Janet, - SHEI, Gan-Ju
: TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D GLUCAN
: TITLE OF INVENTION: SYNTHASE SUBUNITS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.
: STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
: CITY: RAHWAY
: STATE: NJ
: COUNTRY: USA
: ZIP: 07065
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/619,554
: FILING DATE: 01-AUG-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: COPPOLA, JOSEPH A
: REGISTRATION NUMBER: 36,413
: REFERENCE/DOCKET NUMBER: 19104PI
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 732-594-6734
: TELEFAX: 732-594-4720
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1876; amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-619-554-2

Query Match	4.08;	Score 94;	DB 2;	Length 1876;
Best Local Similarity	19.88;	Pred. No. 1.6;		
Matches	79;	Conservative	57;	Mismatches 128;
				Indels 134;
				Gaps 18;

```

0Y      3 DLKTIIDHPNIPMKLII-SGFSIAQFSESYLYTROYOKISETKLPVLEIDEIDEPFK 61
      || : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      392 DLNOLFWEPIEIAIIVLEDGKILIEPL-ERILRLGD-----VWMDVFEFKTYKE 4411
0Y      62 SRNYSRAKAKFSIFGDVYNLAQKLVFTKYDLFPKIMHMAVSL-----NAV 107
      : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      442 TRTMLHTVNTN-----RIMWHMISIFPMYFAVNSPFREYTHNQ 480
0Y      108 LPVFEHMYSTVAQSLCFLG-----LSSLSITVLDPLSYSHVLEKEFGFNKLTVOI 162
      : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      481 QLVNQPOLAAAYKMSACALGTVASLIQIVALCEM-----SEFPRKWAGQHLISREWF 534
0Y      163 TDMIKSLTLAVALICPIILYFLKIFEDKFPFDELYIMVFLPVQIILMTIIPVIMPFN 222
      : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      535 LCIIIFGIL-----GPIIFVAYIKDVIYSTAHAVNAVMEFVAV--ATIIIFSIMPBG 587
0Y      223 KFTPLEDELKKSIESLADRVGFPDLKIFVIDGSRSSHNAVEFGUPTSKRIVLFDTL 282
      || : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      588 LFT-----SYMKX-----TRRVASQT- 605
0Y      283 VNNSSTDEITVLA--HEIGHWOKNHVNMVIF-----SOLHTPLIFS-----LFSIY 322
      : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      606 -----ETVAFAPLHGLDRWM--SYLWVWVFAKXSESYSYFLVLSDROPILRLSTTAM 656
0Y      330 RNTSEYNTGFGFLEKSTGSPVDVPIYTKFPIIIGFMLE 367
      || : : || : : || : : || : : || : : || : : || : : || : : || : :
Db      657 RCTGEY-WMGAVLCK-----YQPRIVIGLVATPIIFL 688

RESULT      8
US-08-477-451-22
Sequence 22, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542

```

INFORMATION FOR SEQ ID NO: 22:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 485 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-477-451-22

Query Match 4.0%; Score 93.5; DB 2; Length 485;
 Best Local Similarity 19.8%; Pred. No. 0.24;
 Matches 93; Conservative 63; Mismatches 159; Indels 155; Gaps 24;

7 IHDHPNIPWKLIIT-----SGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETPH 60
 49 LLSHAKFTMKLIILMTQOMNMFVQ---KFLKTRREISLFPNISKPLKTKIFLSAVT 105
 61 KSRN-----YSRAKAF--SIFG---DYVNL-----AOKLVEIKY--DEPK-- 95
 106 KLAFLKRPYFMYSNPKPFIVSVFRCICVSNLFLIMWFSEKLIKICNIEIKTKRM 165
 96 -IMHNAVSLNVLVPRFHMVSTVAO--SLCFLGLISLSTLVLDPLSTYSHFVLEKFG 152
 166 LVMWONIFVCDWMLIAFESYQIOEFLVCYCEMLPTKTRIRD-P----- 210
 153 FPKLTV-QLMTDMIKSLTAVAGPILYLFKIFDKRPDPLMWYVFLVQILAMT 211
 211 -NKQELTOPKIKGLIMGKILASLGG-----TNLEFTGISLDFSMILVLF----- 256
 212 IIPVIMPENKFTPLEDELKKSIESLADRVGFPLDKI--EVIDGSKSSHNAFTGL 269
 257 ----FLMLMG-----LMDALGKRFNLPMDNINKMAEVLKNG----- 290
 270 PFTSKRIVFQRLVNSNSTDEITAVLAHIGMOKNHIYNAVIFSOHLFFLFSFTSY 329
 291 -----FDSIKNMG-----ALVNGFGSSKSDKTANKMSQVRLMCNLYKVKPF 336
 330 R-----NTSFYNTFGF-----FLEKSTGSEVDPIYTK-----EFPIT--TG 363
 337 HLSRALFQORGNRLLAHYDITTSFLKM---STLDPLFLKIFTTDLAFLACLPFLKRLG 393
 364 FMLFNDLTLPLECAMQF-----VMSLISRTHEYOQADY 396
 394 FFLHRSTISIQYVAIFRNPNIRGSSTSPDALISVMSVSSONPPTFFIY 443

RESULT 9
 US-08-118-270-67
 ; Sequence 67, Application US/08118270
 ; Patent No. 5508384

GENERAL INFORMATION:
 APPLICANT: Murphy, Randall B.
 APPLICANT: Schuster, David I.
 TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 NUMBER OF SEQUENCES: 348
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/118,270
 FILING DATE: 09-SEP-1993
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/943,236
 FILING DATE: 10-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Townsend, Kevin G.
 REGISTRATION NUMBER: 34,033
 REFERENCE/DOCKET NUMBER: MURPHY-2A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633
 INFORMATION FOR SEQ ID NO: 67:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 284 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-118-270-67

Query Match 3.9%; Score 92.5; DB 1; Length 284;
 Best Local Similarity 21.1%; Pred. No. 0.14;
 Matches 62; Conservative 45; Mismatches 96; Indels 91; Gaps 14;

7 IHD-HPNIPWKLIIGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETPHKSRNY 65
 25 LLDHILHPTMYFLNLNSADLCFSSKLQWMO---SQVPSI-----PKIMHNAVSLNVL 108
 66 SRAKAK--FSIFGVYNLAQKLVFIKYDLF----- 108
 67 ACGLAQIYFLEFFGDLGNFL--LVMAAYDRYVAICFPLHYMSIMPKI--ZVSLVVLVSVL 123
 109 PVRFHMVSTVAO--SLCFLGLISLSTLVLDPLSTYSHFVLEKFGNKLTVQMTDMKS 168
 124 -TFH-----AMLTHTLIMARLSFCEDSVIPIHYCDMSTLTKVACSDTHDN 167
 169 LTLAIAIGPILYL-FL-----KIFDKFPDPLMWYVFLVQI 207
 168 ELAIFILGPIYVLPFLLIIVSYARIVSSIFKVPSSQSHKRAFSGHLSVSLFTGTV 227
 208 LAM-----TIIPVFIIM-PMFNKFT-PLEDELKKSIESLADRVGPF 246
 228 IGLVLCPSANNSEVKETWMSIYTVPMLNPFYISLRNDIKDALKIMCKKQIP 281

RESULT 10
 PCT-US93-08528-67
 ; Sequence 67, Application PC/TUS9308528
 ; GENERAL INFORMATION:

APPLICANT: New York University
 TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 NUMBER OF SEQUENCES: 348
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/08528
 FILING DATE: 09-SEP-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/943,236
 FILING DATE: 10-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Townsend, Kevin G.

REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-67

Query Match	3.9%	Score 92.5;	DB 5;	Length 284;
Best Local Similarity	21.1%	Pred. No. 0.14;		
Matches	62;	Conservative	45;	Mismatches 96;
				Indels 91;
				Gaps 14;

```

QY 7 ILD-HPNIPMKLLISGSIAQSFSESVLTQROVKLSSETKPVLVEDEIDDEFHKSRRN 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 25 LLDSHHLPWLFSLNSLFDALCESLKLQNNQ-----SQVPSI-----PY 66

QY 66 SRAKAK---ESISQDYVNLQAKVFIRYDLF-----PRIMHVAVSLNLAVL 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 AGCLAAIQYFFELFEGDLGNFL--LVAAAMYDRVAICPEPLHYSIMSPKI--ZVSLVSLMWL 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 109 PVRFHMYSTVAQSLCPGLLSLSLSTVDLPLSYSHVLEEKGEFNKLTVQMLWTIDIKS 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 -TTH-----AMLTLLMARLSFCEDSVITPHYFCDMSTLTKVACSDHDN 167

QY 169 LTLVAIGGPILYV-FL-----KIFDKPEPTDILMYIMVFLFVQI 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 168 ELAFLIIGGPILVPEFLIIVSYARIYSSIFKVPSSQSIHKAESTCGSHLSVSLFYGV 227

QY 208 LAM-----THIPEIM-PMENKFT-PLDEGLKKSSIESLADVGPP 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 IGLVLCBSANNSVEKVEYMSIYIVAPMLNPFIYSLNRNRDIDKDALEKIMCKQID 281

```

RESULT 11
US-09-091-117-2
Sequence 2, Application US/09091117
Patent No. 6171589
GENERAL INFORMATION:
APPLICANT: The University of Melbourne
TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
TITLE OF INVENTION: Vaccines
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREENEER, WINNER and SULLIVAN P. C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,117
FILING DATE: 12 JUNE 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00803
FILING DATE: 13-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PNT127
FILING DATE: 13-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: WINNER, Ellen P.

```

: TELECOMMUNICATION INFORMATION
: TELEPHONE: +1 303 499 8080
:
: TELEFAX: +1 303 499 8089
:
: INFORMATION FOR SEQ ID NO.: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1030 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
US-09-091-117-2
```

Query Match	3.9%	Score 92.5	DB 4	Length 1030
Best Local Similarity	20.9%	Pred. No. 0.94		
Matches 116, Conservative	88	Mismatches 181	Indels 169	Gaps 31

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/07/582.945
APPLICATION NUMBER: US/07/582.945
FILING DATE: 19901012
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30307/112 PLVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-582-945-2

Query Match 3.9%; Score 92; DB 1; Length 1285;
Best Local Similarity 19.5%; Pred. No. 1.5; Mismatches 133; Indels 126; Gaps 19;
Matches 77; Conservative 59;

QY 110 VRFHNVSTVAOSLCFLGLSLSTLVDPPLSYSHFVEEKGFKLTVQMTIMKSL 169
DB 714 IRFALQDVI-SAPRGVAGAIPEAIDPVKY-----VIED----- 748
QY 170 TLAVAIGPILYLFKIDKPTDPLWYIMVFLVQVILAMTIIVFTMP-----MKNKF 224
DB 749 -----ISVFDKIOTNY-WELPAYESWNEGSNSRLGLRESQSKGMLSKC 793
QY 225 TPLEDQ-----ELKKSIESLADRVGFPDLKIFVIDGSKRSSHNAVFTGLPFTSKRI 276
DB 794 RIENSIVYIGHSEEMFISISPYNOVGPEYL-----YPTFFSM 834
QY 277 V-----LFDTLVNSN-STDEITAVLAHEIGHWOKNHIIVNVIFSQLHT 318
DB 835 LOEVGDLGFEOAFATRNPFNTLVSDRLSLMENTMLTRESFDYTPWDAYIGDINDE--Q 892
QY 319 FLIFSLFTSI-----YRNTSFYN---TFGFLEKSTGSPVDPVITKEPPIITIGMLFND 369
DB 893 FAAMSINERIEKCMNTYRGVAFONSSKSIDFPLNMLT-TFIDNGLTE-----IAISD 943
QY 370 LITPLECAMQFVMSLISRTHEYQA-DAYAKKL-----GYKONLCRALIDLOIK-----NL 418
DB 944 L--PYDIOQETISQFLOGSNEKKTLDAMLFNLDKGDINDGARFKLLQSAKDNNIKRAIGH 1001
QY 419 STMNVDP---YSSYHYSHPTLAERSTALDYSEK 450
DB 1002 SDNSVFPFNPNPKSLYKGNITAEIIEKLDREGOK 1036

RESULT 13
US-08-453-141-2
Sequence 2, Application US/08453141
Patent No. 5885589
GENERAL INFORMATION:
APPLICANT: FOGED, Niels T.
APPLICANT: PETERSEN, Svend
TITLE OF INVENTION: PASTEURELLA VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA

ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/453.141
APPLICATION NUMBER: US/08/453.141
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/293,314
FILING DATE: 22-AUG-1994
APPLICATION NUMBER: US 07/582,945
FILING DATE: 12-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK89/00084
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1995/88
FILING DATE: 04-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40389/102/ARKZO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-141-2

Query Match 3.9%; Score 92; DB 2; Length 1285;
Best Local Similarity 19.5%; Pred. No. 1.5; Mismatches 133; Indels 126; Gaps 19;
Matches 77; Conservative 59;

QY 110 VRFHNVSTVAOSLCFLGLSLSTLVDPPLSYSHFVEEKGFKLTVQMTIMKSL 169
DB 714 IRFALQDVI-SAPRGVAGAIPEAIDPVKY-----VIED----- 748
QY 170 TLAVAIGPILYLFKIDKPTDPLWYIMVFLVQVILAMTIIVFTMP-----MKNKF 224
DB 749 -----ISVFDKIOTNY-WELPAYESWNEGSNSRLGLRESQSKGMLSKC 793
QY 225 TPLEDQ-----ELKKSIESLADRVGFPDLKIFVIDGSKRSSHNAVFTGLPFTSKRI 276
DB 794 RIENSIVYIGHSEEMFISISPYNOVGPEYL-----YPTFFSM 834
QY 277 V-----LFDTLVNSN-STDEITAVLAHEIGHWOKNHIIVNVIFSQLHT 318
DB 835 LOEVGDLGFEOAFATRNPFNTLVSDRLSLMENTMLTRESFDYTPWDAYIGDINDE--Q 892
QY 319 FLIFSLFTSI-----YRNTSFYN---TFGFLEKSTGSPVDPVITKEPPIITIGMLFND 369
DB 893 FAAMSINERIEKCMNTYRGVAFONSSKSIDFPLNMLT-TFIDNGLTE-----IAISD 943
QY 370 LITPLECAMQFVMSLISRTHEYQA-DAYAKKL-----GYKONLCRALIDLOIK-----NL 418
DB 944 L--PYDIOQETISQFLOGSNEKKTLDAMLFNLDKGDINDGARFKLLQSAKDNNIKRAIGH 1001
QY 419 STMNVDP---YSSYHYSHPTLAERSTALDYSEK 450
DB 1002 SDNSVFPFNPNPKSLYKGNITAEIIEKLDREGOK 1036

RESULT 14
US-08-293-314-2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:25:22 ; Search time 277.71 Seconds

(Without alignments)
9100.842 Million cell updates/sec

Title: US-09-165-460a-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaaga.....ggggagataaagaatcaca 2948

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 segs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2711.2	92.0	2850	19	AAV17602	Yeast RCE1 gene en
2	175.2	5.9	936	22	AA58252	Oligonucleotide D1
3	175.2	5.9	936	22	AA58254	Oligonucleotide D1
4	175.2	5.9	936	22	AA58257	Oligonucleotide D1
5	175.2	5.9	936	22	AA58259	Oligonucleotide D2
6	175.2	5.9	936	22	AA58262	Oligonucleotide D2
7	175.2	5.9	936	22	AA58255	Oligonucleotide D1
8	174.8	5.9	936	22	AA58252	Oligonucleotide D1
9	174.8	5.9	936	22	AA58254	Oligonucleotide D1
10	174.8	5.9	936	22	AA58257	Oligonucleotide D1
11	174.8	5.9	936	22	AA58259	Oligonucleotide D2

12	174.8	5.9	936	22	AA58262	Oligonucleotide D2
13	174.8	5.9	936	22	AA58255	Oligonucleotide D1
14	65.6	2.2	244	22	AA58238	Oligonucleotide D1
15	63.6	2.2	244	22	AA58238	Oligonucleotide D1
16	53.6	1.8	6265	20	AA58253	NBP46 (root lectin
17	52.6	1.8	396	22	AA58262	Human ovarian can
18	50.4	1.7	580073	18	AA58262	Myocytoma genital
19	50.2	1.7	612	22	AA58262	Human cervical can
20	49.8	1.7	2229	21	AA58262	CDNA encoding huma
21	49.6	1.7	268	22	AA58262	Human cervical can
22	49.4	1.7	309	22	AA58262	Human cervical can
23	49.2	1.7	2771	22	AA58262	Human cervical can
24	48.8	1.7	310	22	AA58262	Human cervical can
25	48.2	1.6	2046	21	AA58262	Human secreted pro
26	48.2	1.6	6243	20	AA58262	Clostridium specie
27	48.2	1.6	671	22	AA58262	Human secreted pro
28	48.2	1.6	19124	18	AA58262	Plasmodium var-7 g
29	48.2	1.6	19124	21	AA58262	Plasmodium var-7 p
30	47.8	1.6	1117	21	AA58262	Nucleotide sequenc
31	47.6	1.6	4185	22	AA58262	P. falciparum telo
32	47.4	1.6	4237	19	AA58262	Human secreted pro
33	47.2	1.6	2427	11	AA58262	Human pro-urokinas
34	47.2	1.6	2447	19	AA58262	Human secretory pr
35	47.2	1.6	2447	20	AA58262	Human secreted pro
36	46.8	1.6	612	22	AA58262	Human cervical can
37	46.6	1.6	597	22	AA58262	Human cervical can
38	46.4	1.6	731	13	AA58262	Rsal restriction f
39	46.4	1.6	2486	21	AA58262	Nucleotide sequenc
40	46.4	1.6	5139	21	AA58262	Plasmodium falcipa
41	46.4	1.6	240	18	AA58262	Staphylococcus aur
42	45.8	1.6	621	22	AA58262	Human cervical can
43	45.8	1.6	1254	22	AA58262	Human secreted pro
44	45.8	1.6	6033	21	AA58262	Plasmodium falcipa
45	45.8	1.6	9789	17	AA58262	CDNA encoding Plas

ALIGNMENTS

RESULT 1
AAV17602 standard; DNA; 2850 BP.
ID AAV17602:
XX
AC AAV17602:
XX
DT 20-JUL-1998 (first entry)
XX
DE Yeast RCE1 gene encoding Rce1p protein.
XX
KW RCE1 gene; Rce1p; a-factor convertase; CMAX protease;
KW zinc metalloprotease; yeast; prenylation; Ras; inhibitor; cancer;
KW colorectal carcinoma; pancreas carcinoma; leukaemia; therapy; ss.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9805786-A2.
XX
PD 12-FEB-1998.
XX
PF 06-AUG-1997; 97MO-US14777.
XX
PR 30-JUL-1997; 97US-0902774.
PR 07-AUG-1996; 96US-0023491.
XX
(REGC) UNIV CALIFORNIA.
XX
PI Ashby MN, Boyartchuk VL, Rine JD;
DR WPI, 1996-145621/13.
PT Vector comprising nucleic acid coding for Afc1p and Rce1p proteins
PT involved in protein prenylation - useful for screening for
PT inhibitors of proteins, especially of mutated ras proteins involved

PT In cancerous conditions
 XX
 PS Claim 11; Page 48-50; 59pp; English.
 XX
 CC This DNA sequence comprises the yeast RCE1 (Ras and a-factor
 CC convertase) gene that codes for a novel protein (see AAM48302),
 CC designated Rce1p, that mediates the removal of AAX tripeptide from
 CC CAAX proteins such as Ras following prenylation. RCE1 was isolated
 CC using an autocorine arrest, sensilised selection for CAAX proteases.
 CC Another CAAX protease gene, APCI (a-factor convertase) (see AAV17601)
 CC was similarly identified. Vectors comprising the APCI or RCE1
 CC gene, polypeptides encoded by such vectors and recombinant cells
 CC transduced with the vectors are claimed. The novel Apc1p and Rce1p
 CC (see AAM48302) proteins can be used to screen for inhibitors of
 CC prenylation which can be used in the treatment of Ras dependent
 CC cancers such as colorectal and exocrine pancreatic carcinomas or
 CC myeloid leukemia.
 CC (N.B. the translated sequence of this RCE1 gene corresponds to the
 CC Rce1p amino acid sequence provided in the specification only in the
 CC N-terminal region (amino acid residues 1-233)).
 XX
 SQ Sequence 2850 BP; 903 A; 504 C; 518 G; 925 T; 0 other;
 Query Match 92.0%; Score 2711.2; DB 19; Length 2850;
 Best Local Similarity 96.4%; Pred. No. 0;
 Matches 2845; Conservative 0; Mismatches 3; Indels 102; Gaps 3;
 1 tgaactgttgaacaaagaagcctgaacagcatcaaaagcttgcagatgttcca 60
 1 tgaactgttgaacaaagaagcctgaacagcatcaaaagcttgcagatgttcca 60
 61 ttcaaaaattgtgaatacgaacaaagttcaacaagaatttcttgataatgaatt 120
 61 ttcaaaaattgtgaatacgaacaaagttcaacaagaatttcttgataatgaatt 120
 121 atataatgaatagggagacgctgaagcgaagaactgaattatcagaattatcac 180
 121 atataatgaatagggagacgctgaagcgaagaactgaattatcagaattatcac 180
 181 ttccagagcttgaacgatacgaagaagatttgaacttgaattagttgttcacatt 240
 181 ttccagagcttgaacgatacgaagaagatttgaacttgaattagttgttcacatt 240
 241 atccaacattttaaaaaaacagtcagtaaccctaaacaacccaacgaacgcatc 300
 241 atccaacattttaaaaaaacagtcagtaaccctaaacaacccaacgaacgcatc 300
 301 tgaagattcaatcgcacacagtaacgtaagttcatcgtcgccagttcaggtcatt 360
 301 tgaagattcaatcgcacacagtaacgtaagttcatcgtcgccagttcaggtcatt 360
 361 aagcccgtaactcttcacccaagagttatcttctactcagcggtagctctac 420
 361 aagcccgtaactcttcacccaagagttatcttctactcagcggtagctctac 420
 421 ttccctaaattcaagacagcagattgttcataaagtgttgagttgttccttcaagta 480
 421 ttccctaaattcaagacagcagattgttcataaagtgttgagttgttccttcaagta 480
 481 caaacgagacataatgtgaacttgaatttaataaagagacataaagaataatcc 540
 481 caaacgagacataatgtgaacttgaatttaataaagagacataaagaataatcc 540
 541 aagtttgaagaagcgtcctatgttgatttattcgttcgtaagtaattga 600
 541 aagtttgaagaagcgtcctatgttgatttattcgttcgtaagtaattga 600
 601 aaaccactgttgagtagaagaattgatccctatttctgttagcagaagaaata 660
 601 aaaccactgttgagtagaagaattgatccctatttctgttagcagaagaaata 660
 661 aaaaaacattagaaaaaatctcgttacttcttatagatatagatatatgtgttt 720

Db 661 aaaaaacattagaaaaaatctcgttacttcttatagatatagatatatgtgttt 720
 QY 721 gcttatagaagaattattatcgctcccttgatcttccctattatataaattctt 780
 Db 721 gcttatagaagaattattatcgctcccttgatcttccctattatataaattctt 780
 QY 781 taaatgatcttcggcgccctttgtgtctcgtatctttttttttttggaccactg 840
 Db 781 taaatgatcttcggcgccctttgtgtctcgtatctttttttttttggaccactg 840
 QY 841 gatgaaaaaccttgaatgtatttattacaccttatttaagttaactaaatacagagatt 900
 Db 841 gatgaaaaaccttgaatgtatttattacaccttatttaagttaactaaatacagagatt 900
 QY 901 cagagacaaacatagaatttcttcttgcagaagaataaagaagaataatgtgctc 960
 Db 901 cagagacaaacatagaatttcttcttgcagaagaataaagaagaataatgtgctc 960
 QY 961 tgactacgactgtctgcataagagaagacagaacagcaatgtctacattcacaact 1020
 Db 961 tgactacgactgtctgcataagagaagacagaacagcaatgtctacattcacaact 1020
 QY 1021 tctagtctcctatacatctccatactcctatgtctacacgctatatagtcaactcacaac 1080
 Db 1021 tctagtctcctatacatctccatactcctatgtctacacgctatatagtcaactcacaac 1080
 QY 1081 agaagggcttaagaagaatactccgagaagataaactcgcagatcgaataaactaacat 1140
 Db 1081 agaagggcttaagaagaatactccgagaagataaactcgcagatcgaataaactaacat 1140
 QY 1141 tatgttaattccaaccttcttctgtgctcttcttcaacatccaatatactagtaacac 1200
 Db 1141 tatgttaattccaaccttcttctgtgctcttcttcaacatccaatatactagtaacac 1200
 QY 1201 ttcaataataagttccaagaacgcaattttaggtcttagtatacccaagttcaacgc 1260
 Db 1201 ttcaataataagttccaagaacgcaattttaggtcttagtatacccaagttcaacgc 1260
 QY 1261 tgcattgcacaaaccttgcgaacttcagcgaattctgtgaagaactcaacgaatgtgtgc 1320
 Db 1261 tgcattgcacaaaccttgcgaacttcagcgaattctgtgaagaactcaacgaatgtgtgc 1320
 QY 1321 gatgtattgacctatattgtgacccgttttagattttagtattatatacattatataa 1380
 Db 1321 gatgtattgacctatattgtgacccgttttagattttagtattatatacattatataa 1380
 QY 1381 tccaagaagctctacttgaagaattttacacgaattcctgaataatttgaatttcag 1440
 Db 1381 tccaagaagctctacttgaagaattttacacgaattcctgaataatttgaatttcag 1440
 QY 1441 gaatttataatttgcaccaataaactgaggaataatttcaacgtatgaacttgcactac 1500
 Db 1441 gaatttataatttgcaccaataaactgaggaataatttcaacgtatgaacttgcactac 1500
 QY 1501 gtacttaaaccttaacgcagcttgcgaactgaagctatcaagaatttggcaacatc 1560
 Db 1501 gtacttaaaccttaacgcagcttgcgaactgaagctatcaagaatttggcaacatc 1560
 QY 1561 gcttttttttgacttgcgcagcacaacagcttatgtagaattacagaagaagctcat 1620
 Db 1561 gcttttttttgacttgcgcagcacaacagcttatgtagaattacagaagaagctcat 1620
 QY 1621 gacaactgttccactctcgtcgaacaatgcttccaattttatacaacaacttttgg 1680
 Db 1621 gacaactgttccactctcgtcgaacaatgcttccaattttatacaacaacttttgg 1680
 QY 1681 aggtttaacaagaatttgcgaagaacagcggaactatcgtgtgcataacact 1740
 Db 1681 aggtttaacaagaatttgcgaagaacagcggaactatcgtgtgcataacact 1740
 QY 1741 gcatgccttgaacatacatgaggttccctgtccttcaagaattgaattatcattcac 1800
 Db 1741 gcatgccttgaacatacatgaggttccctgtccttcaagaattgaattatcattcac 1800


```

Db 1700 -----ttgcaatacatggygtttaatgctctcaagattgatacttcaac 1750
QY 1801 agtagtagacaagaagaactgagcatttccaaattgg--tctaaactggaataatg 1858
Db 1751 agtagtagacaagaagaactgagcatttccaaattggctctcaactggaataatg 1810
QY 1859 tactgcgactgctgctgctggaataataatccctggaagataccttcaaacctgta 1918
Db 1811 tactgcgactgctgctgctggaataataatccctggaagataccttcaaacctgta 1870
QY 1919 ggaactcctggttataagaataacccctttagccttctttagctactgtatccgtttaa 1978
Db 1871 ggaactcctggttataagaataacccctttagccttctttagctactgtatccgtttaa 1930
QY 1979 atttccatgtactataactcttcttctactactatgtgaattctctatcgagccggg 2038
Db 1931 atttccatgtactataactcttcttctactactatgtgaattctctatcgagccggg 1990
QY 2039 ctttgttgcggaagtggaataaactgagttttgtggtttgttgaagaatttgagg 2098
Db 1991 ctttgttgcggaagtggaataaactgagttttgtggtttgttgaagaatttgagg 2050
QY 2099 actataagtaactataactcttcttctactactataagaagtcgttgcgttgcagtggt 2158
Db 2051 actataagtaactataactcttcttctactactataagaagtcgttgcgttgcagtggt 2110
QY 2159 attgaaagtctcagatctcagaagtagagaagaagtgagcatcctaagtttcgagcttt 2218
Db 2111 attgaaagtctcagatctcagaagtagagaagaagtgagcatcctaagtttcgagcttt 2170
QY 2219 tcttttttaaggttttatttcttctcctcagaattgaagttctttagttttgtttt 2278
Db 2171 tcttttttaaggttttatttcttctcctcagaattgaagttctttagttttgtttt 2230
QY 2279 gttttgtggtttacataatttccaaltccaagaagaattttagctgtcttcttataatgtcca 2338
Db 2231 gttttgtggtttacataatttccaaltccaagaagaattttagctgtcttcttataatgtcca 2290
QY 2339 atagagaatacagagagatgctgtgtactacataaagcgataagcgatcgctagttcaa 2398
Db 2291 atagagaatacagagagatgctgtgtactacataaagcgataagcgatcgctagttcaa 2350
QY 2399 ggggaataacggaagtcctgaaggtttgtgtctgcagacgtcccttgcgaatgaatccgca 2458
Db 2351 ggggaataacggaagtcctgaaggtttgtgtctgcagacgtcccttgcgaatgaatccgca 2410
QY 2459 ggcgtaaaaaacctaaaaagattgtcgaattggttcgaatggaattactatgtatccagat 2518
Db 2411 ggcgtaaaaaacctaaaaagattgtcgaattggttcgaatggaattactatgtatccagat 2470
QY 2519 tagataataaattcgtgtggggaatctagtgggagagatcagtgtctgcgacagatcca 2578
Db 2471 tagataataaattcgtgtggggaatctagtgggagagatcagtgtctgcgacagatcca 2530
QY 2579 gtctgtcgtcaatgcgaagtgacacaccccggttaataacacgataatgacatccaa 2638
Db 2531 gtctgtcgtcaatgcgaagtgacacaccccggttaataacacgataatgacatccaa 2590
QY 2639 ctccgcctagagaactgtgcatgtgaggggttaactcagagatagaatccctccataagacta 2698
Db 2591 ctccgcctagagaactgtgcatgtgaggggttaactcagagatagaatccctccataagacta 2650
QY 2699 aacaaagtaactactatgtatgaataaaaaagtgcttcaactccatccaagaataatg 2758
Db 2651 -----aaatlaatg 2660
QY 2759 ctacgctattaaagaaaaacttatatgtgttccgcgaatcaaaccttcaagtttaagc 2818
Db 2661 ctacgctattaaagaaaaacttatatgtgttccgcgaatcaaaccttcaagtttaagc 2720
QY 2819 ctgataatttctcagagctgtgtacaagatacttacaanaatatacaactaaagcgacaatg 2878
Db 2721 ctgataatttctcagagctgtgtacaagatacttacaanaatatacaactaaagcgacaatg 2780

```

```

QY 2879 gtgaagataatgatcgggaataatgaataaataacgataattgaggaataatggggagata 2938
Db 2781 gtgaagataatgatcgggaataatgaataaataacgataattgaggaataatggggagata 2840
QY 2939 aagaatcaca 2948
Db 2841 aagaatcaca 2850

RESULT 2
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umeek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6: Page 127; 159pp: English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.9%; Score 175.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 2.1e-29;
Matches 9; Conservative 529; Mismatches 252; Indels 0; Gaps 0;

QY 456 ggttagatgctcccttaccagtaacaaagacatgaatgtaacttggaatttaata 515
Db 795 GGGGGGGCGMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWW 736
QY 516 aggaactaaagaaactttaaattccaaagtttgaagaagtcgctatgttagatttatt 575
Db 735 WMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMM 676
QY 576 ggcgtcagtaatgaatgaatttgaaaacacatcctggtgcgtgagatgtatcccta 635
Db 675 WMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMM 616
QY 636 ttctgttagcgaagtgaacaataaataaataacattagaanaaatctctgattcttctt 695

```

```
Db 615 ..... 556
Oy 696 atagatagatatgtatgttgcctatagatgaagatattatgcctctta 755
Db 555 ..... 496
Oy 756 ttcccatattataaatcttttaaatgcatcttctgtctcttctgtctct 815
Db 495 WGN..... 436
Oy 816 gtatttttttttttgacacacgtgagaaaccttgatattataaccttatt 875
Db 435 ..... 376
Oy 876 ttaagttaataaatatcgcgatttcagaaacaaatagaatttcttctgtcaagaa 935
Db 375 ..... 316
Oy 936 aataaacgaaataatgatgcttgcctactgactgctgcatalagagaacagaa 995
Db 315 ..... 256
Oy 996 cagcaatgctcaattctcaacattctagtgctcctatcatctcatctatgtgc 1055
Db 255 ..... 196
Oy 1056 taccgctatgcacttcacacacgaaggctctaaacgagataatcctcgaacagata 1115
Db 195 ..... 136
Oy 1116 aatctgcatacaaaaacttaacatgatctaatctcaaccttttctgtgcctttt 1175
Db 135 ..... 76
Oy 1176 tacaatctcaattctagtcacacttcacataaagttccaagagcattttaggc 1235
Db 75 ..... 16
Oy 1236 tagtatatt 1245
Db 15 ..... 6

RESULT 3
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
AC
XX AAF58254;
AT
XX 24-APR-2001 (first entry)
DE
XX Oligonucleotide D1875.
XX
XX Electrophoretic transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
XX Synthetic.
OS
XX
XX WO200107665-A2.
PN
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-US20476.
PE
XX
XX 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX
XX Umek RM;
PI
XX
XX WPI; 2001-159728/16.
```

```
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SO Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other:

Query Match 5.9%; Score 175.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 2.1e-29;
Matches 9; Conservative 529; Mismatches 252; Indels 0; Gaps 0;

Oy 456 ggttagatgctcccttaccagtaacagagacataatgtaacttgaaattata 515
Db 795 GGGGGGGCGG..... 736
Oy 516 aggaattaaagaaacttaattcgaatttgaaagctgctatgtttagttatt 575
Db 735 ..... 676
Oy 576 ggcctgagtaatgaatttaaaacacatcttgctgctgagagatagataccta 635
Db 675 ..... 616
Oy 636 ttctgttaggcaagtgacaaaataaaacattagaataatctcgtacttctt 695
Db 615 ..... 556
Oy 696 atagatagatatgtatgttgcctatagatgaagatattatgcctctta 755
Db 555 ..... 496
Oy 756 ttcccatattataaatcttttaaatgcatcttctgtctcttctgtctct 815
Db 495 WGN..... 436
Oy 816 gtatttttttttttgacacacgtgagaaaccttgatattataaccttatt 875
Db 435 ..... 376
Oy 876 ttaagttaataaatatcgcgatttcagaaacaaatagaatttcttctgtcaagaa 935
Db 375 ..... 316
Oy 936 aataaacgaaataatgatgcttgcctactgactgctgcatalagagaacagaa 995
Db 315 ..... 256
Oy 996 cagcaatgctcaattctcaacattctagtgctcctatcatctcatctatgtgc 1055
Db 255 ..... 196
Oy 1056 taccgctatgcacttcacacacgaaggctctaaacgagataatcctcgaacagata 1115
Db 195 ..... 136
Oy 1116 aatctgcatacaaaaacttaacatgatctaatctcaaccttttctgtgcctttt 1175
Db 135 ..... 76
Oy 1176 tacaatctcaattctagtcacacttcacataaagttccaagagcattttaggc 1235
Db 75 ..... 16
```

OY 1236 taggtatatt 1245
:: :
Db 15 WWWWMMMMMM 6

RESULT 4
AAFS8257/C
ID AAF58257 standard; DNA: 936 BP.

XX AAF58257:

AC 24-APR-2001 (first entry)

XX Oligonucleotide D1954.

DE Electron-transfer group; ETM: mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

OS WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

XX WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX a single surface

PS Example 6; Page 127; 159pp: English.

XX The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

XX

OY 456 gggttaggattgccttaccagtaacagagacatgaactgaatttaata 515

Db 795 GGGGGGGGCC 736

OY 516 aggacattaaggaacttaattccaaagtttgaagctgcttagttatt 575

Db 735 WWWWMM 676

OY 576 gcgttcagtaataattgaacaacatcttgctgcgaagattgatacccta 635

Db 675 WWWWMM 616

OY 636 ttctgttaggcaagtgcaataaataaataaataaataaataaataaataa 695

Db 615 WWWWMM 556

OY 696 atagatatagatatagtagtctgcttataagataagattatcgccttctga 755

Db 555 WWWWMM 496

OY 756 ttccctattataataaattctttaaagcaatttcgtgctcttctgctct 815

Db 495 WWWWMM 436

OY 816 gtatttttttttttgacacactgagtaaaaccttgatattatattac 875

Db 435 WWWWMM 376

OY 876 ttaagtlactaaatalcgaagattcgaagaaacaaacaaacaaacaaacaa 935

Db 375 WWWWMM 316

OY 936 aataaacgaaataatgtagcttgaactgaactgaactgaactgaactga 995

Db 315 WWWWMM 256

OY 996 cagcaatgctacaattcacaattctagtgctctatcacatccatcctatg 1055

Db 255 WWWWMM 196

OY 1056 taccgctatatgcactcacaccagaaggtctaaacgaataatcgcgaacta 1115

Db 195 WWWWMM 136

OY 1116 aatctgcataacaaactcaattatgcatttccaacatttcttgctctt 1175

Db 135 WWWWMM 76

OY 1176 tacaatcctaattctagtaaccactcacataaagttcaagagcattttag 1235

Db 75 WWWWMM 16

OY 1236 taggtatatt 1245

Db 15 WWWWMMMMMM 6

RESULT 5
AAFS8259/C

ID AAF58259 standard; DNA: 936 BP.

XX AAF58259:

XX 24-APR-2001 (first entry)

XX Oligonucleotide D2004.

DE Electron-transfer group; ETM: mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX a single surface

CC	different redox potentials. The invention is used for electronic
CC	acids each containing an electron-transfer group (ETM) having
CC	a composition comprising two nucleic
PS	Example 6; Page 127; 159pp; English.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
XX	
XX	
PI	umek RM;
XX	
DR	WPI; 2001-159728/16.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PR	17-MAR-2000; 2000US-0190259.
PR	99US-0145695.
PE	26-JUL-1999; 2000WO-US20476.
XX	
PD	01-FEB-2001.
PN	WO200107665-A2.
PN	
OS	Synthetic.
XX	
KM	gene expression; ss.
KW	Electron-transfer group; ETM; mismatch; genotyping;
XX	
DE	Oligonucleotide D1876.
DT	24-APR-2001 (first entry)
XX	
AC	AAFS8255;
XX	
ID	AAF58255 standard; DNA; 938 BP.
RESULT	7

[illegible][illegible]

XX 24-APR-2001 (first entry)
 DT Oligonucleotide D1835.
 DE
 XX Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 KM
 XX Synthetic.
 OS
 XX WO200107665-A2.
 PN
 XX 01-FEB-2001.
 PD
 XX 26-JUL-2000; 2000WO-US20476.
 PF
 XX 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 XX Umek RM;
 PI
 XX WPI: 2001-159728/16.
 DR
 XX
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 CC
 XX
 SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.9%; Score 174.8; DB 22; Length 936;
 Best Local Similarity 0.6%; Pred. No. 2.6e-29;
 Matches 5; Conservative 530; Mismatches 247; Indels 0; Gaps 0;

QY 450 ataaatggttagagatctccttcaagtaacagagacattatgtaattgagat 509
 Db 1
 QY 510 ttaataaggacataaggaaacttaattccaagtttgaagaactgctatgttagt 569
 Db 61
 QY 570 ttatctgcgtcgagtaagtaattgaaccacactctgctgagtaagattgata 629
 Db 121
 QY 630 tccctattctgttagcgaagtgacaaataaaaaacattagaataaattcgttact 689
 Db 181
 QY 690 ttctctatagatagatatagtatgcttctatagatgaagatattatcgctcc 749
 Db 241
 QY 750 ttgatctcctattatataaataattcctttaaatgcaatttctgctcttctgt 809
 Db 301
 QY 810 gctctgatttttttttttctgacacctgagtaagaaaccttgatgatttaacc 869
 Db 361

QY 870 ttatttaagttactaaatatcgagatttcaggaacaaacataagattttctgtc 929
 Db 421
 QY 930 aagaaataaagcaataaattgagcttctgactgactgctgtcatagagaa 989
 Db 481
 QY 990 ccagaacagcaatgtctacaattctcaattcttagtctccttcaaccatctct 1049
 Db 541
 QY 1050 atgctctacgctatattgcaacttccacacaggaaggcttaacagataatcctgaa 1109
 Db 601
 QY 1110 cgattaaatctcgatgcaaaacttaacattatgcttaattccaaccttttctgtgc 1169
 Db 661
 QY 1170 ctttttacaatctcaattatctgtaccacttccacatataagtttcaagagcattt 1229
 Db 721
 QY 1230 ta 1231
 Db 781

RESULT 9
 AAF58254
 ID AAF58254 standard; DNA; 936 BP.
 XX
 AC AAF58254;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D1875.
 DE
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 OS
 PN WO200107665-A2.
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 XX Umek RM;
 PI
 XX WPI: 2001-159728/16.
 DR
 XX
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 CC
 XX
 SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 5.9%; Score 174.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 2.6e-29;
Matches 5; Conservative 530; Mismatches 247; Indels 0; Gaps 0;

```
OY 450 ataaatgggttagatgcttccttaccagtaacagacataatgtgaattggaat 509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1  www..... 60

OY 510 ttaataaggacataaagaaacttaattccaagtttgaagctgctatggtgagt 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61  www..... 120

OY 570 ttatcgctcgagtaatgaattgaaccatcttggcgtagcgaagtgtata 629
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 www..... 180

OY 630 tcccatattctgttagcgaatgacaaataaaacattagaaaaatctcgttact 689
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 www..... 240

OY 690 ttcttataagataagatatgtatggttgcctatagatgaaggtattatcgcgtcc 749
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 www..... 300

OY 750 ttgtatcccatattataaataatctttaaagtcatcttctgtgcttttgtt 809
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 www..... 360

OY 810 gcttcgtatcttttttttggaccacgtgatggaacaccttgatgattatacc 869
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 www..... 420

OY 870 ttattttaagttactaaataatcgagatttcaggaacaaacatagaatttcttctg 929
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 www..... 480

OY 930 aagaaaaataaagaaataatgtatgcttgactgactgctgtctatagagagaa 989
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 www..... 540

OY 990 ccgaacagcaatgctacaattctcaacattctagtctcctatacatccatcc 1049
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 www..... 600

OY 1050 atgtgtacgcgtatagcaacttccaaacagaagggtctaaacgagataatccctgaa 1109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 www..... 660

OY 1110 cgattaaatcgcgtacgaataacttacaattatgctaatcttcttctgtgc 1169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 661 www..... 720

OY 1170 ctttttacaatccaattatctagtaacacttccacataagttccaagacgcat 1229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 721 www..... 780

OY 1230 ta 1231
      : :
Db 781 ww 782
```

RESULT 10

AAFS8257 standard; DNA; 936 BP.

AAFS8257;

24-APR-2001 (first entry)

Oligonucleotide D1954.

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 26-JUL-2000; 2000MO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX a single surface

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic

XX acids each containing an electron-transfer group (ETM) having

XX different redox potentials. The invention is used for electronic

XX detection of nucleic acids, especially of substitutions (mismatches)

XX and single-nucleotide polymorphisms, e.g. for genotyping,

XX monitoring gene expression.

XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.9%; Score 174.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 2.6e-29;
Matches 5; Conservative 530; Mismatches 247; Indels 0; Gaps 0;

```
OY 450 ataaatgggttagatgcttccttaccagtaacagacataatgtgaattggaat 509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1  www..... 60

OY 510 ttaataaggacataaagaaacttaattccaagtttgaagctgctatggtgagt 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61  www..... 120

OY 570 ttatcgctcgagtaatgaattgaaccatcttggcgtagcgaagtgtata 629
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 www..... 180

OY 630 tcccatattctgttagcgaatgacaaataaaacattagaaaaatctcgttact 689
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 www..... 240

OY 690 ttcttataagataagatatgtatggttgcctatagatgaaggtattatcgcgtcc 749
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 www..... 300

OY 750 ttgtatcccatattataaataatctttaaagtcatcttctgtgcttttgtt 809
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 www..... 360

OY 810 gcttcgtatcttttttttggaccacgtgatggaacaccttgatgattatacc 869
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 www..... 420

OY 870 ttattttaagttactaaataatcgagatttcaggaacaaacatagaatttcttctg 929
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 www..... 480

OY 930 aagaaaaataaagaaataatgtatgcttgactgactgctgtctatagagagaa 989
```

Query Match	5.98;	Score 174.8;	DB 22;	Length 936;
Best Local Similarity	0.68;	Pred. No. 2.6e-29;		
Matches	5;	Conservative 530;	Mismatches 247;	Indels 0;
			Gaps	0;

Qy	450	ataaagtggttagatgtgcctctttaaagtaacagagacatlaatlgtaacttggaa	509
Dp	1	60
Qy	510	ttaaataaggacatlaagaaacttlaattccaaagtttgaaagctgcatagttgta	569
Dp	61	120
Qy	570	tttatcgctcgcgtatgaatgaatattgtaaaacacattcctggcgtacggaattgta	629
Dp	121	180
Qy	630	ttccattctcgttaagcaatgacaaataaanaaacatttagaanaaaattccgttact	689
Dp	181	240
Qy	690	tttcctatagatagatataatgattgttcttatagatgaaggtattatccgtctc	749
Dp	241	300
Qy	750	tttgattccctattatataaaattcctttaaatgcatttccgtgacctttgt	809
Dp	301	360
Qy	810	gctctgatttttttttttggccactgataagaaaccttgatatttacc	869
Dp	361	420
Qy	870	tttatcttaagttactaaatcatcgagatttcagaaacaaatagaatttcttgc	929
Dp	421	480
Qy	930	aagaaataaagcaataatgatgcttgaactgactgctgcataagagaa	989
Dp	481	540
Qy	990	ccagaaacgaatgctacaattccaaacttctagttgctcatatccatccatccct	1045
Dp	541	600
Qy	1050	atggtcaccgcctatagcaacttcacaccgaagggctctaaagagataatcctgaa	1105
Dp	601	660
Qy	1110	cgattaaatccgatgcaaaaacttaacaattgctaattccaaacttttctgtgc	1165
Dp	661	720
Qy	1170	ctttttacaatctcaattatctagtaaccatccatataaagttccaagagcatct	1225
Dp	721	780
Qy	1230	ta 1231	
Dp	781	

RESULT	12
AAF58262	
ID	AAF58262 standard; DNA; 936 BP.
XX	
AC	AAF58262;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D2007.
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	


```

Oy 510 ttaataaggacataaggaaacttaattccaagtttgaagcgtcctatgttgaagt 569
Db 61 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 120
Oy 570 tttattgcttcgagtaagttaatttgaaccactctggcgtaagagatgta 629
Db 121 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 180
Oy 630 tcccaatttcgttaggaagtgacaaataaaaaacattagaaaaatttcgttact 689
Db 181 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 240
Oy 690 tttcctatagataagataatgtagtctgctatagaaagtattatcgcgtcc 749
Db 241 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 300
Oy 750 ttgtatccctattatataaattctttaaattcatttctgtgctcttctt 809
Db 301 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 360
Oy 810 gcttcgtatttttttttttttgacacgtgatgaaaccttgatgatttattacc 869
Db 361 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 420
Oy 870 ttattttaagttactaaatalcgaagattcaggaacaaacataagaatttcttctc 929
Db 421 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 480
Oy 930 aagaataataaagaataaataatgtagcttgactgactgctgctatagaagaa 989
Db 481 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 540
Oy 990 ccagaacagcaatgctacaatttccaacatttctagtcctctatcacatcatalcct 1049
Db 541 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 600
Oy 1050 atgtgctacgcgtatagcaatcacaacagaaggcttaacgagataatcctcgaa 1109
Db 601 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 660
Oy 1110 cgaataatcgcgtacaaacttaacattatgctaatttccaaccttcttctgtgc 1169
Db 661 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 720
Oy 1170 cttttttaaactcaattatctagtaaccacttccatatagaatttcaaggaacgatttt 1229
Db 721 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 780
Oy 1230 ta 1231
Db 781 ww 782

```

```

RESULT 14
AAF58238/c
ID AAF58238 standard; DNA: 244 BP.
XX
AC AAF58238;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1250:D1102.
XX
KM Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN MO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.

```

```

XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Uniek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 4; page 120; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 2.28; Score 65.6; DB 22; Length 244;
Best Local Similarity 6.48; Pred. No. 2.1e-05;
Matches 14; Conservative 145; Mismatches 59; Indels 0; Gaps 0;

Oy 737 attatcgagccttctgtatccctattatataaattctttaaattcatttctctg 796
Db 222 ATTATGCTGCTTTTAAACAMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 163
Oy 797 gtgctcttctgtctctgtatttttttttttgagaccagtgatgaaaccttga 856
Db 162 WMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 103
Oy 857 tgatttataccttattttaaattcctaataatcagagattcaggaacaaactag 916
Db 102 WMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 43
Oy 917 aatttcttgcagaagaaataaacaagaaataatcg 954
Db 42 WMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 5

RESULT 15
AAF58238
ID AAF58238 standard; DNA: 244 BP.
XX
AC AAF58238;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1250:D1102.
XX
KM Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN MO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX

```

PI Umek RM;
 XX
 DR WPI; 2001-159728/16.
 XX

PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX

PS Example 4; Page 120; 159pp; English.
 XX

CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX

SO Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 2.2%; Score 63.6; DB 22; Length 244;
 Best Local Similarity 2.9%; Pred.No. 5.8e-05;
 Matches 6; Conservative 148; Mismatches 52; Indels 0; Gaps 0;

OY 754 tttccctatttaataaattctttaaatgcatlcttggtgctctttgtgctt 813
 Db 1 tctacagmwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 60
 OY 814 ctgtatttttttttttgacacactgagtgaacaccttgatgatttatacctta 873
 Db 61 www 120
 OY 874 tttaagttaactaaataatcgagatttcagagacaaacatagaatttcttgcaga 933
 Db 121 www 180
 OY 934 aaataaaacgaataaatgatgct 959
 Db 181 www 206

Search completed: February 12, 2002, 13:25:38
 Job time: 8176 sec

THIS PAGE BLANK (LJPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:01:41 ; Search time 2973.29 Seconds

(without alignments)
10634.378 Million cell updates/sec

Title: US-09-165-460A-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaaag.....ggggagagataaagatcaca 2948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estlom:*
5: em_estlpl:*
6: em_estlba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gD_estl:*
11: gD_est2:*
12: gD_hic:*
13: gD_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_liv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428	14.5	994	13	CNS066GJ
2	161.6	5.5	519	13	AZ926837
3	140	4.7	1045	13	CNS06POM
4	113	3.8	1134	13	CNS06NMS
5	86.2	2.9	497	13	AZ929776
6	85.6	2.9	933	13	CNS06PEO
7	83	2.8	1111	13	CNS06LMU
8	79.4	2.7	1101	13	CNS0039G
9	75.6	2.6	1125	10	AL547503
10	75.2	2.6	1101	13	CNS00EVL
11	70.8	2.4	996	13	CNS00FUF
12	66.8	2.3	1101	13	CNS001FB

C 13	66.8	2.3	1101	13	CNS003BD	AL064091	Drosophila
C 14	66.6	2.3	619	10	AL514935	AL514935	AL514935
C 15	66.4	2.3	840	13	CNS010TW	AL099422	Drosophila
C 16	66.4	2.2	759	13	CNS06QXV	AL411257	77 end of
C 17	65.4	2.2	1043	13	CNS0145P	AL103735	Drosophila
C 18	65.4	2.2	1225	13	CNS0161D	AL106171	Drosophila
C 19	64.6	2.2	404	10	AL514087	AL514087	AL514087
C 20	64.6	2.2	811	13	AZ538053	AZ538053	EMTGA06TF
C 21	64.4	2.2	1101	13	CNS0127P	AL102007	Drosophila
C 22	64.2	2.2	329	10	AL513719	AL513719	AL513719
C 23	63.8	2.2	1101	13	CNS00EFA	AL069119	Drosophila
C 24	63.4	2.2	955	13	BH136744	BH136744	EMTGA02TF
C 25	63.4	2.2	1101	13	CNS0039G	AL063921	Drosophila
C 26	63.4	2.2	1101	13	CNS00EWF	AL069847	Drosophila
C 27	63	2.1	366	10	AL514543	AL514543	AL514543
C 28	63	2.1	1190	13	CNS020M7	AL20620M7	Tetradon
C 29	62.8	2.1	966	13	CNS0055C	AL061991	Drosophila
C 30	62.6	2.1	335	10	AL513597	AL513597	AL513597
C 31	62.6	2.1	1001	13	CNS0155H	AL105023	Drosophila
C 32	62.4	2.1	1101	13	CNS00L72	AL078714	Drosophila
C 33	62.2	2.1	1101	13	CNS00K55	AL077453	Drosophila
C 34	62	2.1	401	10	AL515191	AL515191	AL515191
C 35	62	2.1	1092	13	CNS020K7	AL175636	Tetradon
C 36	62	2.1	1101	13	CNS00FYG	AL071206	Drosophila
C 37	61.8	2.1	425	10	AL514791	AL514791	AL514791
C 38	61.8	2.1	1101	13	CNS0021J	AL061936	Drosophila
C 39	61.6	2.1	748	13	CNS03FUD	AL241563	Tetradon
C 40	61.2	2.1	867	13	CNS00CX5	AL060052	Drosophila
C 41	61	2.1	625	13	CNS036A2	AL229763	Tetradon
C 42	61	2.1	625	13	CNS036A2	AL229763	Tetradon
C 43	60.8	2.1	758	10	AV756150	AV756150	AV756150
C 44	60.8	2.1	1101	13	CNS00E07	AL069440	Drosophila
C 45	60.4	2.0	353	10	AL515235	AL515235	AL515235

ALIGNMENTS

RESULT 1	CNS066GJ	994 bp	DNA	GSS	17-JUN-2001
LOCUS	T7 end of clone AS0A002A09 of library AS0A from strain CLIB 533				
DEFINITION	of Saccharomyces bayanus, genomic survey sequence.				
ACCESSION	AL397313				
VERSION	AL397313.1	GI:12149936			
KEYWORDS	GSS.				
SOURCE	Saccharomyces bayanus.				
ORGANISM	Saccharomyces bayanus				
REFERENCE	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
AUTHORS	Elkaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; 1 (bases 1 to 994)				
TITLE	Bon,E., Nevegilise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.				
JOURNAL	Genomic exploration of the hemiascomycetous yeasts: 5.				
MEDLINE	Saccharomyces bayanus var. ivarium				
REFERENCE	PEBS Lett. 487 (1), 37-41 (2000)				
AUTHORS	2 (bases 1 to 994)				
TITLE	Souciot,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Nevegilise,C., Ozler-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissbach,J.				
JOURNAL	Genomic exploration of the hemiascomycetous yeasts: 1. A set of				
MEDLINE	yeast species for molecular evolution studies				
REFERENCE	FEBS Lett. 487 (1), 3-12 (2000)				
AUTHORS	20584711				
TITLE	3 (bases 1 to 994)				
JOURNAL	Genoscope.				
MEDLINE	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage.				
AUTHORS	Direct Submission				
TITLE	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage.				
JOURNAL	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage.				
	2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :				

[illegible]

ORIGIN	Query Match	4.7%	Score 140;	DB 13;	Length 1045;
	Best Local Similarity	55.4%	Pred. No. 2,5e-14;		
	Matches 297;	Conservative	0;	Mismatches 230;	Indels 9; Gaps 14;
QY	65	aaatctgtgataagaacacgaatccaagaattcttcgtgaattaaagatlat	124		
Db	518	AAAATCAAGATATATATTAATTCAGATTCTACGAAATGCTGATTAATTCATCATTTAT	577		
QY	125	aatatgaatagggagaccgtaggcaccaagaagactgaaattacagatttacttca	184		
Db	578	AATGTAATTCGTAGATTATTATATCTGCGCAATTAATAATTCACGATTCATCTAGT	637		
QY	185	cagctggttaacagatctgaaagccttgcgaactggaagttagtgcgtcacactacc	244		
Db	638	CAAAATTCATATGACATTGAAAGTTTACCTAAATTTAAAGTTCAACCATTAATCT	697		
QY	245	aaccttcttaaaaaacagtcagtlacccctaaacaacactcaagcacgcatlctgag	304		
Db	698	GATATTTCAGAAACATTAATGGAACCTTTAAAGAAAGATGATTCATCGTTA-----	751		
QY	305	gattcaataatgcacacacagtgtaagtgatgttcacatgcgcgcagttcagcgtcattacg	364		
Db	752	--ATGATTACACCAACCAATTCACAGTAGTACGACACCAAAAGAGAACATCGCAAG	808		
QY	365	ccagtaactcttcacccaagtagtatttaccatagcggtagctgcgtactcc	424		
Db	809	ATTGTCATGCTAAAGTGCTTGTTGGAAACATGCTACATCTCTAGAAACCTCTATATCG	868		
QY	425	ctgaaattacagacagatgtgtcataaagtggttagtgattgcttccttaccagtaaca	484		
Db	869	AATAGTGTACTGACCAATTAATTCGCAATGCGAAACAAAGGTAATCAATATGAG	928		
QY	485	cgaagcatctatgtgaacttgatattataagaacattaaagaacttaattccagt	544		
Db	929	AGATAGATCAATGTTAATTTAGCAATTCATTAATACATCTTGAAGACATTTGGTACCACGT	988		
QY	545	tttgaagctgcctcatgttgtagtatttatttcggttcgagtaatgaattga	600		
Db	989	TTTGAAACTTGTTATGTTGTGATTTTATTGTGTAGTAGTGAACATTTAAATTTCA	1044		
RESULT	4				
CNS06NM5/c	CNS06NM5	1134 bp	DNA	GSS	04-JUL-2001
LOCUS	T7 end of clone AU00AA015C11 of library AU00A from strain CBS 3082				
DEFINITION	T7 end of clone AU00AA015C11 of library AU00A from strain CBS 3082				
ACCESSION	AL006947				
VERSION	AL006947.1	GI:12171645			
KEYWORDS	GSS.				
SOURCE	Saccharomyces kluyveri.				
ORGANISM	Saccharomyces kluyveri				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
AUTHORS	Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Attienave,F.,				
TITLE	1 (bases 1 to 1134)				
JOURNAL	Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Attienave,F.,				
MEDLINE	Galliardin,C., and Casaregola,S.				
REFERENCE	Genomic exploration of the hemiascomycetous yeasts: 9.				
AUTHORS	Saccharomyces kluyveri				
	FEBS Lett. 487 (1), 56-60 (2000)				
	20584719				
	2 (bases 1 to 1134)				
	Soucieu,T.L., Aigle,M., Attienave,F., Blandin,G.,				
	Boitout-Pukhara,M., Bon,E., Brothier,P., Casaregola,S.,				
	de-Montigny,J., Dujon,B., Durieux,P., Lepingle,A., Llorente,B.,				
	Maupertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,				
	Saurin,W., Tekala,F., Toffano-Nicchoe,C., Wesolowski-Louvel,M.,				
	Wincker,P. and Weissenbach,J.				
	Genomic exploration of the hemiascomycetous yeasts: 1. A set of				
	yeast species for molecular evolution studies				
	FEBS Lett. 487 (1), 3-12 (2000)				

Db	82	AATAGCTGTTTCATTTTTTTCCACAGTCGTATACACTCAATATTTGGTGTTTCACAA	23
Oy	1693	gttttattcgtgaagaacagg	1713
Db	22	CTTATCTTCATCAGACAGG	2
RESULT	6		
CNS06PBO/c			
LOCUS			
DEFINITION	CNS06PBO	933 bp	DNA
ACCESSION	T7 end of clone AV0AA012H11 of library AV0AA from strain CBS 379 of	GSS	05-JUL-2001
KEYWORDS	Saccharomyces exiguus, genomic survey sequence.		
ORGANISM	AlA09138		
SOURCE	GI:12176372		
REFERENCE	GSS.		
AUTHORS	Saccharomyces exiguus.		
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
TITLE	Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Attiguenave,F.,		
JOURNAL	Gallardin,C. and Casaregola,S.		
MEDLINE	Genomic exploration of the hemiascomycetous yeasts : 6.		
REFERENCE	Saccharomyces exiguus		
AUTHORS	FEB5 Lett. 487 (1), 42-46 (2000)		
	20584716		
	2 (bases 1 to 933)		
	Souci�t,J.L., Aigle,M., Attiguenave,F., Blandin,G.,		
	Bojotin-Fukuhara,M., Bon,E., Brotlier,P., Casaregola,S.,		
	de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,		
	Malpertuy,A., Neuveglise,C., Ozler-Kalogeropoulos,O., Potier,S.,		
	Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,		
	Wincker,P. and Weissenbach,J.		
	Genomic exploration of the hemiascomycetous yeasts : 1. A set of		
	yeast species for molecular evolution studies		
	FEB5 Lett. 487 (1), 3-12 (2000)		
	20584711		
	3 (bases 1 to 933)		
	Genoscope.		
	Direct Submission		
	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,		
	2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :		
	seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
	This GSS is part of a random genomic sequencing program of thirteen		
	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces		
	exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,		
	Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces		
	lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia		
	angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,		
	Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to		
	5 kb were prepared and both extremities were sequenced. See		
	keywords for description of this sequence and for the sequence of		
	the other extremity of this insert.		
	Location/Qualifiers		
	1..933		
	/organism="Saccharomyces exiguus"		
	/strain="CBS 379"		
	/db_xref="taxon:34358"		
	/clone="AV0AA012H11"		
	/clone_lib="AV0AA"		
	/note="end : T7"		
	complement(1..11..>856)		
	/note="similar to Saccharomyces cerevisiae ORF YML11w [
	strong similarity to ubiquitination protein Bulp j]"		
	/evidence-not_experimental		
	complement(137..>932)		
	/note="similar to Saccharomyces cerevisiae ORF YMR25c [
	Bul1 : ubiquitination pathway protein]		
	1 putative frameshift(s)"		
	/evidence-not_experimental		
BASE COUNT	312 a	141 c	130 g
ORIGIN		347 t	3 others

Query Match 2.9% Score 85.6 DB 13 Length 933;
Best Local Similarity 57.5% Pred. No. 4e-05;
Matches 154; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy	443	attgtcctaataatgggttagagattgcctcccttaccagtaacaagagacatlaatgtgaac	502
Db	346	ATTGGTAAGGTGGTGTGTACCAAGTTGGAACACAGAAATTGAAAGATTGTCTTCGCAGAT	287
Qy	503	tttgaatttaataagagacattaagaaacttaattccaaattttgaaactgcctatgt	562
Db	286	TTAGAAATGACATCGGATATTGAGAAACCTTAATTCACATTTTGAAAGATTCGTATAGT	227
Qy	583	tttagtgatttttcgcttcagagtaatgatttaatttgaataacccctctggtaggaag	622
Db	226	TGTCCTTCTCTATGCATTTAGAGTTAATTAATTGATATGTCACATGGATCAGCCAGT	167
Qy	623	atgatacccatattctcgttaggcaagtgcacaaataaaaaacatagaaaaaatctt	682
Db	166	CTTGATATTCCAAATCAATGTTAAAAAGATGAATATTGATGTCGATACGATTAACGAGAT	107
Qy	683	cgttaattctcttatagatatagata	710
Db	106	AGATACATATTTTATTTTAAATTTTA	79

RESULT 7
CNS06LNU/C
LOCUS
DEFINITION T3 end of clone XAT0AA002B08 of library XAT0AA from strain CBS 4311
ACCESSION AL404740
VERSION AL404740.1 GI:12166586
KEYWORDS
SOURCE
ORGANISM
SACCHAROMYCES servazzii.
SACCHAROMYCES servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 1111)
Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
Artiguenave,F., Wincker,P. and Gaillardin,C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)
20584717
2 (bases 1 to 1111)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bollotin-Pukhara,M., Bon,E., Brotier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Majpetu,A., Neuveglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
3 (bases 1 to 1111)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
serfe@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the extremity of this insert.
location/Qualifiers

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL071063

VERSION AL071063.1 GI:4951105

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 996)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila melanogaster BAC library](http://www.fruitfly.org/TheBDGP/Drosophila%20melanogaster%20BAC%20library) was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

location/Qualifiers

source 1. 996

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_1lb="RPCI-98"

/clone="BACR31021"

/note="end : TET3"

BASE COUNT 383 a 164 c 81 g 171 t 197 others

ORIGIN

Query Match 2.4%; Score 70.8; DB 13; Length 996;

Best Local Similarity 37.4%; Pred. No. 0.012;

Matches 151; Conservative 71; Mismatches 176; Indels 6; Gaps 1;

620 aagattgataccctatttcgttagcgaagtcgcaaaataaataacattagaataat 679

945 WAMATTWMTAATTTBTBTBGGSSGSSGSSGSGRRRAAAAAAAMWMTAT 886

680 tctgtaaccttctatagatagatatatgctgctatagatagaagatct 739

885 TTTTWTAAWMTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 826

740 taccgctccttgcctatccatataataaattccttaaatgcatcttcgtg 799

825 WTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 772

800 cctcttgcctcctgcatcttcttcttcttgcgacacgaggaacacctgatga 859

771 TTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 712

860 tttatcacccttcttcttcttcttcttcttcttcttcttcttcttcttctt 919

711 TTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 652

920 tttcttgcgaagaataaataacgaataatgcatgcttgcctgactgctgctc 979

651 TTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 592

980 atagagagaccgagacgcaatgctcaaatcttcaaatcttcc 1023

Db 591 TMAAATATKADTWAMWTTWATKTTTTTTTATTTATTTATTTTWT 548

RESULT 12

CNS001FB 1101 bp DNA GSS 03-JUN-1999

LOCUS BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL060732

VERSION AL060732.1 GI:4939397

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila melanogaster BAC library](http://www.fruitfly.org/TheBDGP/Drosophila%20melanogaster%20BAC%20library) was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

location/Qualifiers

source 1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_1lb="RPCI-98"

/clone="BACR04A23"

/note="end : TET3"

BASE COUNT 288 a 110 c 103 g 491 t 109 others

ORIGIN

Query Match 2.3%; Score 66.8; DB 13; Length 1101;

Best Local Similarity 38.8%; Pred. No. 0.056;

Matches 165; Conservative 55; Mismatches 199; Indels 6; Gaps 1;

532 ttaattccaagtttgaagcctcctatgcttgaagtttattgctgctgagtaagat 591

675 TTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 734

592 taatttgaagaacatcttgcgtagaagaatgataccctatcttcttgcgtaag 651

735 WAAATTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 794

652 gacaaataaataacatlaaataaattcctgcttcttcttcttcttcttctt 711

795 ATTWATTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 848

712 gtagtgcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 771

849 ATWATTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 908

772 aaattctttaaataatgcatcttcttgcgcttcttgccttcttcttcttctt 831

909 AMATTAAWTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 968

[illegible]

LOCUS	AL514935	619 bp	mRNA	EST	13-FEB-2001
DEFINITION	AL514935 LTL_NFL006_P12		Homo sapiens	CDNA clone	Cl0BB0122D08.3
ACCESSION	AL514935				
VERSION	AL514935.1	GI:12778428			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Li, W. B., Gruber, C., Jessae, J. and Polares, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Genoscope				

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secrete@genoscope.cns.fr, Web : www.genoscope.cns.fr

FEATURES

source

location/Qualifiers

1. 619

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Cl0B80122D08"
/clone_1bp="LTI_NFL006_pl2"

was primed with a NotI-oligo(dT)₁₈ primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filang@lifetech.com URL : <http://fulllength.invitrogen.com>

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 15:33:56 : Search time 91.73 Seconds
(without alignments)
7278.504 Million cell updates/sec

Title: US-09-165-460a-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaag.....ggggagagataaagaatcacac 2948

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.2	1.6	6243	2	US-09-056-075-1
C 2	48	1.6	19124	2	US-08-487-826B-13
C 3	47.8	1.6	1117	4	US-09-247-373B-33
C 4	47	1.6	2447	2	US-09-014-969-14
C 5	46.4	1.6	731	1	US-08-451-405A-2
C 6	46	1.6	240	1	US-08-628-417-6
C 7	45	1.5	3012	2	US-08-475-427-5
C 8	45	1.5	3012	2	US-07-842-165-5
C 9	44.6	1.5	636	4	US-08-998-416-1137
C 10	44.6	1.5	1483	4	US-09-262-749-1
C 11	44.2	1.5	5852	1	US-07-867-106-2
C 12	43.6	1.5	1230	3	US-09-022-669-3
C 13	43.6	1.5	19124	2	US-08-487-826B-13
C 14	43	1.5	1553	3	US-09-022-669-1
C 15	42.8	1.5	7218	1	US-08-232-463-14
C 16	42.6	1.4	837	4	US-08-998-416-288
C 17	42.6	1.4	3138	1	US-07-867-106-4
C 18	42.4	1.4	578	4	US-09-020-956-104
C 19	42.4	1.4	578	4	US-09-030-607-104
C 20	42	1.4	2110	4	US-09-419-459-1
C 21	42	1.4	5852	1	US-07-867-106-2
C 22	41.6	1.4	4376	1	US-08-119-125A-1
C 23	41.6	1.4	6744	1	US-08-119-125A-2
C 24	41.2	1.4	1474	4	US-08-821-994-64
C 25	40.8	1.4	2230	3	US-08-844-188-34
C 26	40.4	1.4	8920	2	US-08-446-855A-1
C 27	40.4	1.4	8920	4	US-09-150-741-1

28	40.2	1.4	658	4	US-08-998-416-595	Sequence 595, App
C 29	40.2	1.4	686	4	US-08-953-326-21	Sequence 21, Appl
C 30	40.2	1.4	1641	1	US-08-300-803A-8	Sequence 8, Appl
C 31	40.2	1.4	4970	1	US-08-764-100-14	Sequence 14, Appl
C 32	40.2	1.4	4970	1	US-08-764-100-14	Sequence 20, Appl
C 33	40	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl
C 34	39.8	1.4	688	6	5498694-3	Patent No. 5498694
C 35	39.8	1.4	960	3	US-09-248-335-57	Sequence 57, Appl
C 36	39.8	1.4	1602	1	US-08-530-850-3	Sequence 3, Appl
C 37	39.8	1.4	1602	3	US-08-888-429A-3	Sequence 3, Appl
C 38	39.8	1.4	1602	4	US-09-149-879-3	Sequence 3, Appl
C 39	39.4	1.3	1243	2	US-08-687-080-53	Sequence 53, Appl
C 40	39.4	1.3	1582	3	US-08-545-196B-10	Sequence 10, Appl
C 41	39.4	1.3	1582	3	US-08-545-196B-12	Sequence 12, Appl
C 42	39.4	1.3	5893	1	US-08-592-126-54	Sequence 54, Appl
C 43	39.4	1.3	5893	2	US-08-687-080-44	Sequence 44, Appl
C 44	39.2	1.3	926	3	US-08-945-994-8	Sequence 8, Appl
C 45	39.2	1.3	2058	2	US-08-749-391-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-056-075-1/c
Sequence 1, Application US/09056075
Patent No. 595368
GENERAL INFORMATION:
APPLICANT: Johnson, Eric A.
APPLICANT: Bradshaw, Marile
APPLICANT: Rood, Julian
TITLE OF INVENTION: Expression System for Clostridium
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
City: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,075
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95238
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 3770..4013
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orig) from
US-09-056-075-1."
Query Match 1.6%; Score 48.2; DB 2; Length 6243;

RESULT 4
US-09-014-969-14/c
: Sequence 14, Application US/09014969


```

: Patent No.5965397
: PATENT INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallee, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: AGOSTILINO, Michael J.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/014,969
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sprunger, Suzanne A.
: REGISTRATION NUMBER: 41,323
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8284
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2447 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-09-014-969-14

Query Match 1.6%; Score 47; DB 2; Length 2447;
Best Local Similarity 49.4%; Pred. No. 0.026;
Matches 119; Conservative 1; Mismatches 121; Indels 0; Gaps 0;

QY 679 ttctcgtacttcttatagatagatatgtaagtgttgcttatagatgaaggat 738
Db 2446 ttttttttttttttttttttttttttttttttttttttttttttttttttt 2387
QY 739 ttatcggcgtcttgtatccctcatattataaaaaatcctttaaaagcaatttcgt 798
Db 2386 ttttttttttttttttttttttttttttttttttttttttttttttttttt 2327
QY 799 gctcttggtgctctgtatattttttttttttggacacgtgatgtaaacccttgatg 858
Db 2326 ttttttttttttttttttttttttttttttttttttttttttttttttttt 2267
QY 859 atttataccttattttaagttaacaaatatcgagaattccaagaaacaacatagaa 918
Db 2266 ttttttttttttttttttttttttttttttttttttttttttttttttttt 2207
QY 919 t c 919
Db 2206 T 2206

RESULT 5
US-08-451-405A-2/c
: Sequence 2, Application US/08451405A
```

```

Patent No. 5736358
GENERAL INFORMATION:
APPLICANT: FASEL, NICOLAS JOSEPH
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE WEBB LAW FIRM
STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
CITY: PITTSBURGH
STATE: PENNSYLVANIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISK
COMPUTER: Midwest Micro 486-50
OPERATING SYSTEM: DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,405A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,273
FILING DATE: 15-JAN-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 731
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: UNKNOWN
US-08-451-405A-2

Query Match      1.6%   Score 46.4; DB 1; Length 731;
Best Local Similarity    53.3%; Pred. No. 0.023;
Matches          98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Db 648 aagtcgcaaaatlaaaaaacattagaataaaatcttcgttactttcttaagatatagat 707
      ||| ||||| ||||| ||| |||| |
Db 488 AATTAAAAAAATTTAAAAAAATTTATTTAAAATAATTAATTAATAATTAATAAAATTTAAA 429
Qy 708 ataigtatggttgcttatagatgaaggatattatcgcgtcccttgtaatlcccatlct 767
Db 428 TTTTAAACCTTGTAAGATGAATGAGGTGTTGTTAAAGAAGTGTGTTGGCAATTAAGATA 369
Qy 768 aataaaattcctttaaaaalgcattcttcggtgcccctttgttgcttcgtatctttct 827
      ||||| ||||| ||||| ||| |||| |
Db 368 TATAATATATATTTTAAAAAATTTCTTTAAATACCTTTTTTTTGTGATTTTATTTTTTTT 309
Qy 828 tttt 831
Db 308 TTTT 305

RESULT
US-08-628-417-6/c
Sequence 6, Application US/08628417
Patent No. 5627054
GENERAL INFORMATION:
APPLICANT: GILLESPIE, DAVID
TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
TITLE OF INVENTION: POLYMERASE CHAIN REACTION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
ADDRESS: DEFENSE COMMAND
STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-CC)
CITY: ABERDEEN PROVING GROUND
STATE: MARYLAND
COUNTRY: USA
ZIP: 21010-5423
```


Db 670 TAAATAATCTGAACACATTATGAAGACATTGGCTTGTGATATGTAGGAAA 614

RESULT 8

```

US-07-842-165-5/c
Sequence 5, Application US/07842165
Patent No. 5932698
GENERAL INFORMATION:
APPLICANT: DUBOIS, Michel
APPLICANT: GRISON, Rene
APPLICANT: LEGUAY, Jean-Jacques
APPLICANT: PIGNARD, Annie
APPLICANT: TOPPAN, Alain
TITLE OF INVENTION: Recombinant gene coding for a protein
TITLE OF INVENTION: having endochitinase activity or for a precursor thereof
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: King Street Station, Suite 500, 1800 Diagonal
STREET: Road, PO Box 299
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0239
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/842,165
FILING DATE: 19920501
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/348
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3012 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: intron
LOCATION: 2384..2462
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: intron
LOCATION: 2617..2697
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: join(1942..2383, 2463..2616, 2698..3007)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1942..2166
FEATURE:
NAME/KEY: mat_peptide
LOCATION: join(2167..2383, 2463..2616, 2698..3007, 2698)
US-07-842-165-5

```

OY 135 gatttgctatatacaacttaaaacccaagagctcctaactggaattttaccat 1414
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 970 GAATTTCTTAAATGAATAAATAACTCATATAAAAATTACTAGAGAATTTGAAATTTA 911

OY 1415 gaattccgatatttggagtcttcagaatatctatactgcacaabaacttagaanaa 1474
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 910 TGAATTTCGAATTTATATGTATGCATTCACACCTCCAAAGCTTCTTAAATACGGATTAAGAATTT 851

OY 1475 ttctaacgcgtcaatgctctttgacctagctacttaaactaataccgcatctcgaaactaac 1534
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 850 TGACAATAATTAATTACTTTATTTGTGATTCCTGATTCGTAGTGATCPTAGTGGGAGTCGCCCGTC 791

OY 1535 tatcacagttatttggcaacaacatcgctcttttttggacctgcgcacacacatgct 1594
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 790 GCCCAATATTATTTTGTGTTGTGATGGAACCTTAAATTTCTACGTGAACCTCAAAAATGTGTGT 731

OY 1595 tatgagaattatcacggaaggctccacagacaacacglttcatctctgcgcgacaacatgctc 1654
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 730 AATGTGCAATTTAAATTCGAATTTTGTCGCTAAATTCATATAACATGACGCCCAAGTTTGACA 671

OY 1655 caaatcttatcacacaacacttttggagggttaccaaagtcttgtatcttgtaagaaa 1711
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 TAAATTAATCTGACACATTAATTAATGAAGACATTTGGTGTCGTTGATATSTAGAAAAA 614

RESULT

```

US-08-998-416-1137
Sequence 1137, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NO. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GCCL1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8688
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

```


RESULT 12
US-09-022-669-3
Sequence 3 Application US/09022669-3
Patent No. 610717
GENERAL INFORMATION:
APPLICANT: KILLY, KRISTINE
APPLICANT: SOUTHAN, CHRISTOPHER
APPLICANT: KNAB, ANNE
TITLE OF INVENTION: Human RCEL
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: RAYNER & PRESTIA

RESULT 13
 US-08-487-826B-13
 : Sequence 13, Application US/08487826B
 : Patent No. 5993827
 : GENERAL INFORMATION:
 : APPLICANT: Sim, Kim L.
 : APPLICANT: Chluis, Chetan
 : APPLICANT: Miller, Louis H.
 : APPLICANT: Peterson, David S.
 : APPLICANT: Su, Xia-zhaun
 : APPLICANT: Wellens, Thomas E.
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEIN
 : NUMBER OF SEQUENCES: 45
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Knobbe Martens Olson & Bear
 : STREET: 620 Newport Center drive 16th floor

THIS PAGE BLANK (1-2PTO)

PT in cancerous conditions
XX
PS Claim 15; Page 52-53; 59pp; English.
XX
CC This polypeptide comprises yeast Rce1p (Ras and a-factor converting
CC enzyme), a novel zinc metalloprotease that mediates the removal of
CC AAx tripeptide from CAAX proteins following prenylation. The yeast
CC Rce1 gene (see AAV17601) was isolated using an autoclave arrest
CC selection method. A second novel CAAX protease, Afcep1 (a-factor
CC converting enzyme), is also claimed (see AAM48301). Vectors
CC comprising the Rce1 or Afcep1 gene (see AAV17601), polypeptides encoded
CC by such vectors and recombinant cells transduced with the vectors
CC are claimed. The novel Afcep1 and Rce1p proteins can be used to
CC screen for inhibitors of prenylation which can be used in the
CC treatment of Ras dependent cancers such as colorectal and exocrine
CC pancreatic carcinomas or myeloid leukaemia. A heat shock assay for
CC assessing ras activity is also provided.
CC (N.B. this amino acid sequence of Rce1p only corresponds to the
CC translated sequence of the Rce1 gene provided in the specification
CC in the N-terminal region (amino acid residues 1-233)).
XX
SQ Sequence 315 AA:
Query Match 99.1%; Score 1636; DB 19; Length 315;
Best Local Similarity 99.4%; Pred. No. 1.1e-178;
Matches 313; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DB 1 MLOSTFLVLYLXISIVLPYATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVPIQ 60
DB 1 mlgstflvlylxisivlpdyatsqegskrdnprikstmkqltlnlslnllvplfq 60
QY 61 SOLSSSTSHISFKDAFLGILIPGYAALPNPWOFSQFVNDLTKVAMLLTYCGPYLDF 120
DB 61 sqsstshisfkdaflgilyipyaaalpnpwqfsqfvdltkcvamlltlycgpvldf 120
QY 121 VLYHLNPKSSILEDFFHEFLNIMSFNRFAPITTEIFTSMLTTYLNLIPHSQLSYQ 180
DB 121 vlyhlnpkssiledffheflnimsfnrfapitteesifysmlltlylnlphsqlsyq 180
QY 181 QLFWOPSLFEGLAHANHAEOLOGSMTVSIILTTFCQILYTLFGSLTFEVVPRGNG 240
DB 181 qlfwopslfeglahanahayeglgqegsmtvsiilttcfqilytlfgsltfvtrvrgn 240
QY 241 LMCCIILHALCNIMGFPGPSRLNLHFTVVDKAKGRISKLVSIWKKCYFALLVGLISLKD 300
DB 241 lwcclilhalcnimgfpgpsrlnlhftvvdckagrisklvsiwkkcyfallvlglsikd 300
QY 301 TLQTLVGTGPGYRITL 315
DB 301 tlqtlvgtpgyritl 315
RESULT 2
AAV55811 standard; Protein: 314 AA.
XX
AC AAV55811;
XX
DT 29-FEB-2000 (first entry)
XX
DE Yeast RCE1 (farnesyl-directed endopeptidase) sequence.
XX
KM RCE1; farnesyl-directed endopeptidase; oncogenic ras; cancer; yeast;
XX cell proliferative disease; cellular localization; tumor.
XX
OS Saccharomycetes cerevisiae.
XX
PD 02-DEC-1999;
XX
PF 19-APR-1999; 99WO-0508594.

XX
PR 22-MAY-1998; 98US-0086421.
XX
PA (ONYX-) ONYX PHARM INC.
XX
PI Choi Y, North AK, Martin GA, Bollag G;
XX
DR WPI; 2000-097123/08.
XX
PT New polypeptides, useful to identify specific modulators that inhibit
PT maturation of oncogenes such as ras, e.g. for treating cancer and other
PT cell proliferative diseases
PS
XX Disclosure; Fig 4; 31pp; English.
CC
CC The invention provides isolated RCE1 (farnesyl-directed endopeptidase)
CC polypeptides. The RCE1 polypeptides are used to identify its specific
CC modulators, i.e. agents that inhibit maturation of oncogenic ras, so are
CC potentially useful for treatment of cancer and other cell proliferative
CC diseases. Fragments of RCE1 are used to raise specific antibodies and
CC other RCE1 ligands such as nucleic acid aptamers which are used to detect
CC RCE1 for diagnostic or research purposes, including cellular localization
CC and/or distribution of the RCE1, also as therapeutic modulators and for
CC identifying RCE1-binding agents by immunoprecipitation. Nucleic acid that
CC encodes RCE1 is used: for recombinant expression of RCE1; as a source of
CC probes and primers for detecting, quantifying and isolating RCE1 e.g. for
CC diagnosis, including detecting mutations that may be associated with
CC cancers and benign tumors; as a source of antisense or ribozyme
CC therapeutics for regulating RCE1 expression; and/or as size marker. The
CC present sequence represents the yeast RCE1 amino acid sequence.
XX
SQ Sequence 314 AA:
Query Match 98.2%; Score 1621.5; DB 21; Length 314;
Best Local Similarity 98.7%; Pred. No. 4.8e-177;
Matches 311; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
DB 1 MLOSTFLVLYLXISIVLPYATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVPIQ 60
DB 1 mlgstflvlylxisivlpdyatsqegskrdnprikstmkqltlnlslnllvplfq 60
QY 61 SOLSSSTSHISFKDAFLGILIPGYAALPNPWOFSQFVNDLTKVAMLLTYCGPYLDF 120
DB 61 sqsstshisfkdaflgilyipyaaalpnpwqfsqfvdltkcvamlltlycgpvldf 120
QY 121 VLYHLNPKSSILEDFFHEFLNIMSFNRFAPITTEIFTSMLTTYLNLIPHSQLSYQ 180
DB 121 vlyhlnpkssiledffheflnimsfnrfapitteesifysmlltlylnlphsqlsyq 179
QY 181 QLFWOPSLFEGLAHANHAEOLOGSMTVSIILTTFCQILYTLFGSLTFEVVPRGNG 240
DB 181 qlfwopslfeglahanahayeglgqegsmtvsiilttcfqilytlfgsltfvtrvrgn 239
QY 241 LMCCIILHALCNIMGFPGPSRLNLHFTVVDKAKGRISKLVSIWKKCYFALLVGLISLKD 300
DB 240 lwcclilhalcnimgfpgpsrlnlhftvvdckagrisklvsiwkkcyfallvlglsikd 299
QY 301 TLQTLVGTGPGYRITL 315
DB 300 tlqtlvgtpgyritl 314
RESULT 3
AAV55810 standard; Protein: 329 AA.
XX
AC AAV55810;
XX
DT 29-FEB-2000 (first entry)
XX
DE Mouse RCE1 (farnesyl-directed endopeptidase) sequence.
XX

KW RCE1: farnesyl-directed endopeptidase; oncogenic ras; cancer.
 KM cell proliferative disease; cellular localization; tumor.
 OS Mus sp.
 MN MO9961628-A2.
 PM 02-DEC-1999.
 PD 19-APR-1999; 99WO-US08594.
 PR 22-MAY-1998; 98US-0086421.
 XX (ONIX-); ONIX PHARM INC.
 PA Choi Y, North AK, Martin GA, Bollag G;
 PI WPI: 2000-097123/08.
 DR N-PSDB: AA239729.
 XX
 PT New polypeptides, useful to identify specific modulators that inhibit
 PT maturation of oncogenes such as ras, e.g. for treating cancer and other
 PT cell proliferative diseases
 PS Claim 8; Fig 3; 31pp. English.
 XX
 XX The invention provides isolated RCE1 (farnesyl-directed endopeptidase)
 CC polypeptides. The RCE1 polypeptides are used to identify its specific
 CC modulators, i.e. agents that inhibit maturation of oncogenic ras, so are
 CC potentially useful for treatment of cancer and other cell proliferative
 CC diseases. Fragments of RCE1 are used to raise specific antibodies and
 CC other RCE1 ligands such as nucleic acid aptamers which are used to detect
 CC RCE1 for diagnostic or research purposes, including cellular localization
 CC and/or distribution of the RCE1, also as therapeutic modulators and for
 CC identifying RCE1-binding agents by immunoprecipitation. Nucleic acid that
 CC encodes RCE1 is used: for recombinant expression of RCE1; as a source of
 CC probes and primers for detecting, quantifying and isolating RCE1 e.g. for
 CC diagnosis, including detecting mutations that may be associated with
 CC cancers and benign tumors; as a source of antisense or ribozyme
 CC therapeutics for regulating RCE1 expression; and/or as size marker. The
 CC present sequence represents the mouse RCE1 amino acid sequence.
 XX
 XX Sequence 329 AA:
 SO
 Query Match 14.1%; Score 233.5; DB 21; Length 329;
 Best Local Similarity 30.9%; Pred. No. 4.1e-18;
 Matches 86; Conservative 41; Mismatches 80; Indels 71; Gaps 16.
 QY 4 ESTFLVLLYISIVLPIYATSQPEGSKRDNPRTIKSRWOKTLMISNLFLVFLQSG 62
 Db 40 fscf---slacysgslsywkwse--lprdhpaivkrr--stsvlvselsplcvllwre 91
 QY 63 LSTST---TSHSFKRAFLGLGIIPEYVALLPMPQFSQFVKDITGVAMLLT--LYGCV 117
 Db 92 lTgIqpgtslItlm-gfIlegIfp--aal-----lalllmIIfIgpI 131
 QY 118 LDFV-----LYHLNPKRSITLDFYHEFLNWS-----FRNRFAPITEELF 159
 Db 132 mqlsmdcedpdltdglkvvlaprs-----warcltdmrwltnqvlaplteeIv 178
 QY 160 YTSMLTTTYLNLIPHSQLSYQOLFQWPSLFFGCIANAHNAHVEQLQSGSMYTVSILLTTCPO 219
 Db 179 fracmIpm---laprtglg-pavfctcp-lffgvahfnhleglrlfrgsavgslftfsaatq 233
 QY 220 ILTYTLFEGGLKRFVVRPTGCNLMCCITILHALCNMGFP 257
 Db 234 fsytavfgayatafItlr-ghIlgpvlchsfcmymgIfp 270
 RESULT 4
 ID AAW86010 standard; Protein: 329 AA.

```

XX AAW86010;
XX
XX 29-MAR-1999 (first entry)
XX
DE Mouse CAXX processing enzyme RCE1 homologue mRceIp.
XX
XX RCE1; mRceIp; mRceIp; CAXX processing enzyme; mouse; tumour; cancer;
XX KW therapy; diagnosis; Ras protein; endoprotease.
XX
XX Mus sp.
XX
XX WO9854333-A2.
XX
XX 03-DEC-1998.
XX
PD 02-JUN-1998; 98WO-US11415.
XX
XX 14-JUN-1997; 97US-0052389.
XX PR 02-JUN-1997; 97US-0047369.
XX
XX (ACAC-) ACACIA BIOSCIENCES INC.
XX
PI Ashby MN, Dimster-Denk DG, Phillips JW;
XX
XX WPI: 1999-059843/05.
XX DR N-PSDB: AAV80323-24.
XX
XX New DNA encoding mammalian CAXX-processing enzymes - used e.g. to
PT treat CAXX-protein mediated diseases such as cancers and tumours
PT associated with mutant Ras
XX
XX Claim 15; Fig 2B; 98pp; English.
XX
CC This is the amino acid sequence of mRceIp, a murine functional
CC homologue of the yeast prenylation-dependent CAXX endoproteinase
CC RcepI that contributes to the processing of a-factor and the yeast
CC Ras protein. This mammalian homologue represents a potential
CC target to block the oncogenic action of mutant Ras protein in
CC tumours or, more generally, to modulate the activity of prenylated
CC peripheral membrane proteins. The mRceIp amino acid sequence was
CC deduced from the nucleotide sequence of isolated cDNA and genomic
CC clones (see AAW80323-24). Host cells transformed with mammalian
CC CAXX processing enzyme DNA (see AAV80322-25) can be used to produce
CC recombinant polypeptides (see AAW86009-12) used for in vitro
CC screening of inhibitors and to raise antibodies. The inhibitors
CC are used to treat CAXX-protein mediated diseases, especially
CC cancers and tumours associated with abnormal Ras activity.
CC Antibodies are used to screen for expression of CAXX processing
CC proteins, for affinity purification and in immunoassays to
CC determine levels of CAXX processing proteins or their subcellular
CC localisation and to confirm interaction with candidate binding
CC proteins.
XX
XX Sequence 329 AA;
XX
XX
XX Query Match 14.0%; Score 230.5; DB 20; Length 329;
XX Best Local Similarity 30.6%, Pred. No. 9,1e-18;
XX Matches 85; Conservative 40; Mismatches 82; Indels 71; Gaps 16;
XX
XX 4 FSTFVLVYISISIVLPYATSQPEGSRDNDRTRTKSRMOKTITIMLSNLF-LVPFLQSQ 62
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX slacysgslsyvwmkse--lprdhpaavlkr--stsvlvsslaplc|clwre 91
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 63 LSSST---TSHSFKDAFLGLGIIPETVALPWPWQFSQPWKDITGVAMLLT--LYCGPV 117
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX ltlglqpgtslltm-gflleglfpa-a|lp-----llltm|flfp| 131
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 118 LDFV-----LYHLNPKSSSLDEFHYHFLINWS-----FNRFPAPIEERF 159
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 132 mq|smddpcdltldg|kvvlapis-----varc|tdmrwtlnqvaplteeiv 178

```

QY 160 YTSMLLTYYLNLIPHSOLSYOQLFMPQSLFPGLAHAAHAYEOLQEGSMYTVSILLTTCFQ 219
 Db 179 fracmlpm---lapctglg-pavftcp-lffgvahfhhleqlfrqsgvgnfltsaafq 233
 QY 220 ILTYTLFGGLTKFVFRGNNLCCIIHALCNIMGFP 257
 Db 234 fsyavfaycaflfirt-ghlpgvichsfcmymgfp 270

RESULT 5

AAW86009 standard; Protein: 293 AA.

AC AAW86009;

DT 29-MAR-1999 (first entry)

DE Human CAAX processing enzyme Rce1 homologue hRce1p.

KM Rce1; Rce1p; hRce1p; CAAX processing enzyme; human; tumour;

XX cancer; therapy; diagnosis; Ras protein; endoproteolase.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Region 172..175

FT /note= "histidine-rich sequence, possible involved in substrate binding and/or catalysis"

PN W09854333-A2.

PD 03-DEC-1998.

PF 02-JUN-1998; 98WC-US11415.

PR 14-JUL-1997; 97US-0052389.

PR 02-JUN-1997; 97US-0047369.

PA (ACAC-) ACACIA BIOSCIENCES INC.

PI Ashby MN, Dimster-Denk DG, Phillips JW;

DR WPI; 1999-059843/05.

DR N-PSDB; AAV80322.

PT New DNA encoding mammalian CAAX-processing enzymes - used e.g. to treat CAAX-protein mediated diseases such as cancers and tumours associated with mutant Ras

PS Clalm 15; Fig 2A; 98pp; English.

XX This is the amino acid sequence of hRce1p, a human functional
 CC homologue of the yeast prenylation-dependent CAAX endoprotease
 CC Rce1p that contributes to the processing of a-factor and the yeast
 CC Ras protein. This mammalian homologue represents a potential
 CC target to block the oncogenic action of mutant Ras protein in
 CC tumours or, more generally, to modulate the activity of prenylated
 CC peripheral membrane proteins. The hRce1p amino acid sequence was
 CC deduced from the nucleotide sequence of cDNA clones (see AAV80322)
 CC isolated from a 9-wk foetus cDNA library. Host cells transfected
 CC with mammalian CAAX processing enzyme DNA sequences (see AAV80322-25)
 CC can be used to produce recombinant polypeptides (see AAW86009-12)
 CC used for in vitro screening of inhibitors and to raise antibodies.
 CC The inhibitors are used to treat CAAX-protein mediated diseases,
 CC especially cancers and tumours associated with abnormal Ras
 CC activity. Antibodies are used to screen for expression of
 CC CAAX processing proteins, for affinity purification and in
 CC immunosays to determine levels of CAAX processing proteins or
 CC their subcellular localisation and to confirm interaction with
 CC candidate binding proteins.

XX Sequence 293 AA;

Query Match 13.9%; Score 229.5; DB 20; Length 293;
 Best Local Similarity 30.6%; Pred. No. 1e-17;
 Matches 83; Conservative 39; Mismatches 82; Indels 67; Gaps 15;

QY 11 LYSISVPLPLVATSOEGRKNDPRTRKSMQKRLTLMISNLF-LVPFIQSLSST--- 66

Db 7 lslacsvgsilyvwkse--lprdhavlkrrf--tsvlvsvslaplcvlwreltqlpg 62

QY 67 TSHISFKDAFLGLGIIIGYAAALPNPWOFSQFVADLTKCVMALLT--LYGSPVDFV--- 121

Db 63 tslltlm-gtrlegifpa--allp-----liltmllflplmqldmcd 102

QY 122 -----LYHLNPKSSILEDYHEFLINIS-----FRNFAPITEIEFYTSMLT 166

Db 103 pcdladglkvvlaprs-----warcldmrvinqyaplteelvitracmlp 149

QY 167 TYLNLIPHSOLSYOQLFMPQSLFPGLAHAAHAYEOLQEGSMYTVSILLTTCFQILYTLT 226

Db 150 m---lapcmglg-pavftcp-lffgvahfhhleqlfrqsgvgnfltsaafqsfyavf 204

QY 227 GGLTKFVFRGNNLCCIIHALCNIMGFP 257

Db 205 gaytaflfirt-ghlpgvichsfcmymgfp 234

RESULT 6

AAV26897 standard; Protein: 329 AA.

AC AAV26897;

DT 07-DEC-1999 (first entry)

DE Human farnesylated--protein converting enzyme 2 protein.

KM Human; farnesylated-protein converting enzyme; Face; homology; Rce-1;

KM Saccharomyces cerevisiae; proteolysis; prenylation; metalloprotease;

XX inhibitor; enzyme; degradation; Ras; diagnosis; cancer.

OS Homo sapiens.

PN W09935275-A1.

PD 15-JUL-1999.

PF 08-JAN-1999; 99WC-ES00001.

PR 08-JAN-1998; 98ES-0000016.

PA (UYOV-) UNIV OVIEDO.

PI Perez Freije JM, Velasco Cotarelo G, Martin Pendas A, Blay Albors P;

PI Balbin Felechosa M, Lopez Olin C;

DR WPI; 1999-430402/36.

DR N-PSDB; AAV83494.

PT New human sequences encoding enzymes that degrade prenylated proteins, useful in diagnosis and treatment of cancers

PS Disclosure; Page 19-20; 28pp; Spanish.

XX This sequence represents the human farnesylated-protein converting
 CC enzyme 2 (Face-2) which has homology to the Saccharomyces cerevisiae
 CC Rce-1 protein. Rce-1 is an enzyme involved in the proteolytic processing
 CC of prenylated proteins, especially factor-a or Ras. This protein and the
 CC Face-1 protein (AAV26896) can be used to design specific inhibitors of
 CC enzymes involved in degradation of prenylated proteins, particularly Ras
 CC and related proteins and for treating disease mediated by Ras and
 CC related proteins and/or the genes encoding them, specifically for
 CC diagnosis and treatment of cancers.

PR	18-MAY-1999	9905-01347168
PR	19-MAY-1999	9905-01347124
PR	20-MAY-1999	9905-01351424
PR	21-MAY-1999	9905-01355523
PR	24-MAY-1999	9905-01356529
PR	25-MAY-1999	9905-01363621
PR	27-MAY-1999	9905-01365921
PR	28-MAY-1999	9905-01367682
PR	01-JUN-1999	9905-01373222
PR	03-JUN-1999	9905-01375228
PR	04-JUN-1999	9905-01375702
PR	07-JUN-1999	9905-01377124
PR	08-JUN-1999	9905-01380694
PR	10-JUN-1999	9905-01385470
PR	10-JUN-1999	9905-01388447
PR	14-JUN-1999	9905-01394152
PR	16-JUN-1999	9905-01394952
PR	16-JUN-1999	9905-01394952
PR	17-JUN-1999	9905-01394952
PR	18-JUN-1999	9905-01394546
PR	18-JUN-1999	9905-01394545
PR	18-JUN-1999	9905-01394546
PR	18-JUN-1999	9905-01394546
PR	18-JUN-1999	9905-01394545
PR	18-JUN-1999	9905-01397653
PR	18-JUN-1999	9905-01396177
PR	21-JUN-1999	9905-01401991
PR	22-JUN-1999	9905-01398699
PR	23-JUN-1999	9905-01403554
PR	23-JUN-1999	9905-01403554
PR	24-JUN-1999	9905-01406955
PR	28-JUN-1999	9905-01408823
PR	29-JUN-1999	9905-01409191
PR	30-JUN-1999	9905-01412877
PR	01-JUL-1999	9905-01414542
PR	01-JUL-1999	9905-01412544
PR	02-JUL-1999	9905-01420250
PR	06-JUL-1999	9905-01423950
PR	08-JUL-1999	9905-01428603
PR	09-JUL-1999	9905-01429270
PR	12-JUL-1999	9905-01435342
PR	13-JUL-1999	9905-01435342
PR	15-JUL-1999	9905-01440605
PR	16-JUL-1999	9905-01440865
PR	16-JUL-1999	9905-01443321
PR	19-JUL-1999	9905-01443321
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443334
PR	20-JUL-1999	9905-01443352
PR	20-JUL-1999	9905-01443632
PR	20-JUL-1999	9905-01446884
PR	21-JUL-1999	9905-01448114
PR	21-JUL-1999	9905-01450586
PR	21-JUL-1999	9905-01450588
PR	23-JUL-1999	9905-01452524
PR	23-JUL-1999	9905-01455214
PR	26-JUL-1999	9905-01452766
PR	27-JUL-1999	9905-01455513
PR	27-JUL-1999	9905-01455513

	PR	27-JUL-1999;	99US-01459519.		
	PR	28-JUL-1999;	99US-0145951.		
	PR	02-AUG-1999;	99US-0146386.		
	PR	02-AUG-1999;	99US-0146389.		
	PR	03-AUG-1999;	99US-01472038.		
	PR	04-AUG-1999;	99US-0147204.		
	PR	05-AUG-1999;	99US-0147302.		
	PR	05-AUG-1999;	99US-0147192.		
	PR	06-AUG-1999;	99US-0147260.		
	PR	06-AUG-1999;	99US-0147416.		
	PR	09-AUG-1999;	99US-0147493.		
	PR	10-AUG-1999;	99US-0147935.		
	PR	11-AUG-1999;	99US-0148171.		
	PR	12-AUG-1999;	99US-0148341.		
	PR	13-AUG-1999;	99US-0148565.		
	PR	13-AUG-1999;	99US-0148684.		
	PR	16-AUG-1999;	99US-0149368.		
	PR	17-AUG-1999;	99US-0149175.		
	PR	18-AUG-1999;	99US-0149426.		
	PR	20-AUG-1999;	99US-0149722.		
	PR	20-AUG-1999;	99US-0149723.		
	PR	20-AUG-1999;	99US-0149929.		
	PR	23-AUG-1999;	99US-0149902.		
	PR	23-AUG-1999;	99US-0149930.		
	PR	25-AUG-1999;	99US-0150566.		
	PR	26-AUG-1999;	99US-0150884.		
	PR	27-AUG-1999;	99US-0151065.		
	PR	27-AUG-1999;	99US-0151066.		
	PR	27-AUG-1999;	99US-0151060.		
	PR	30-AUG-1999;	99US-0151303.		
	PR	31-AUG-1999;	99US-0151438.		
	PR	01-SEP-1999;	99US-0151930.		
	PR	07-SEP-1999;	99US-0152363.		
	PR	10-SEP-1999;	99US-0153070.		
	PR	13-SEP-1999;	99US-0153758.		
	PR	15-SEP-1999;	99US-0154018.		
	PR	16-SEP-1999;	99US-0154039.		
	PR	20-SEP-1999;	99US-0154779.		
	PR	22-SEP-1999;	99US-0155139.		
	PR	23-SEP-1999;	99US-0155486.		
	PR	24-SEP-1999;	99US-015659.		
	PR	28-SEP-1999;	99US-0156596.		
	PR	29-SEP-1999;	99US-0157117.		
	PR	04-OCT-1999;	99US-0157753.		
	PR	05-OCT-1999;	99US-0157865.		
	PR	06-OCT-1999;	99US-0158029.		
	PR	07-OCT-1999;	99US-0158232.		
	PR	08-OCT-1999;	99US-0158369.		
	PR	12-OCT-1999;	99US-0159293.		
	PR	13-OCT-1999;	99US-0159294.		
	PR	13-OCT-1999;	99US-0159295.		
	PR	14-OCT-1999;	99US-0159329.		
	PR	14-OCT-1999;	99US-0159330.		
	PR	14-OCT-1999;	99US-0159331.		
	PR	14-OCT-1999;	99US-0159637.		
	PR	14-OCT-1999;	99US-0159638.		
	PR	18-OCT-1999;	99US-0159584.		
	PR	21-OCT-1999;	99US-0160741.		
	PR	21-OCT-1999;	99US-0160767.		
	PR	21-OCT-1999;	99US-0160768.		
	PR	21-OCT-1999;	99US-0160770.		
	PR	21-OCT-1999;	99US-0160814.		
	PR	21-OCT-1999;	99US-0160815.		
	PR	22-OCT-1999;	99US-0160980.		
	PR	22-OCT-1999;	99US-0160981.		
	PR	22-OCT-1999;	99US-0160989.		
	PR	25-OCT-1999;	99US-0161404.		
	PR	25-OCT-1999;	99US-0161405.		
	PR	25-OCT-1999;	99US-0161406.		
	PR	26-OCT-1999;	99US-0161359.		
	PR	26-OCT-1999;	99US-0161360.		
	PR	26-OCT-1999;	99US-0161361.		
	PR	28-OCT-1999;	99US-0161920.		
	PR	28-OCT-1999;	99US-0161922.		
	PR	28-OCT-1999;	99US-0161993.		
	PR	29-OCT-1999;	99US-0162142.		
	Query Match	5.9%; Score 97.5; DB 21; Length 461;			
	: Best Local Similarity	22.1%; Pred. No. 0.024;			
	Matches	44; Conservative	30; Mismatches	86; Indels	39; Gaps
OY	70	ISFKDAFLGAGITIPGYAALPNPQESOPKVD-----LTCKVAMLLTTCGPVLDPLV	122		
Dd	277	isfplallgtvgrnsgapnpvrktpdrpekkwyltprsvslm---g9llpf--	330		
OY	123	YHLNPKSILEDVEHFLNMSFRNLFPATFEIRFTYSMLTTYNLI-PHSQLSYOQ	181		
Dd	331	-----gsliemyftfstwmkyvyvgfmllvtlvalivveclvgtyfllnaen	383		
OY	182	LFWOPSLEFGIAHAHAHAYEOLQEGSMTTVSILTTFCQILTYTLFGL--TKPEVVRTGG	239		
Dd	384	yhwqwtstffsa-----stavvylysiyyvyvkcmgsffqtstfyf---gy	427		
OY	240	NLWCCIIHALCNIMGRPG	258		
Dd	428	tmmfcldglgcagavylg	446.		
RESULT	13				
ID	AAG45639	standard; Protein; 490 AA.			
XX	AAG45639;				
DF	18-OCT-2000	(first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 57322.				
KM					

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137752.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144844.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147250.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147307.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154739.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156556.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

Query Match	5.9%	Score 97.5	DB 21	Length 490
Best Local Similarity	22.1%	Pred. No. 0.026		
Matches 44	Conservative 30	Mismatches 86	Indels 39	Gaps 7
QY	70	ISFNKAFGLGIGIPCYVALPMPQFSGPVKD-----LTKCVAMLLFLYCGPVIDEVL	122	
Db	306	ISFIALIIVGVVGRWSGAPNRCVKKTIPIRIPEKKWYIIPSVSLM---GGLIPF--	359	
QY	123	YHLNPKRSSILEDYFHEFLNWSFRNFIFAPYTEEIFYFSMLTYLNLI-PSOLSXOO	181	
Db	360	-----GSIFTEMVFVFTSFNRYKVYVYGFMLIVFVILVTCVTLVGTGFLINAEN	412	
QY	182	LFWQPSLFEGLANAHNAHAYEQLOEGSMYTVSILLTTCFQILVYTLFEGL--TKFVFPVRTGG	239	
Db	413	YHWQVTSFSSA-----STAVYVLYSLIYYVYVTKMGGFFQTSFYF--GY	456	
QY	240	NLMCCIIILHALCNIMGFPG	258	
Db	457	tmnfclgllcgavgylg	475	
RESULT 14				
AAG21346				
ID	AAG21346 standard; Protein; 496 AA.			
XX				
AC	AAG21346;			
XX				
DT	17-OCT-2000 (first entry)			
XX				
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 23871.			
XX				
KW	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
KW	termination sequence.			
XX				
OS	Arabidopsis thaliana.			
XX				
PN	EPI033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
PE	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999;	99US-0121825.		
PR	09-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0123548.		
PR	23-MAR-1999;	99US-0125788.		
PR	23-MAR-1999;	99US-0126264.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	08-APR-1999;	99US-0128714.		
PR	16-APR-1999;	99US-0128845.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	23-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	30-APR-1999;	99US-0132407.		

PR	04-MAY-1999;	9905-01323485;
PR	05-MAY-1999;	9905-01323486;
PR	06-MAY-1999;	9905-01323487;
PR	07-MAY-1999;	9905-01323488;
PR	11-MAY-1999;	9905-0134256;
PR	14-MAY-1999;	9905-0134218;
PR	14-MAY-1999;	9905-0134219;
PR	14-MAY-1999;	9905-0134221;
PR	14-MAY-1999;	9905-0134370;
PR	18-MAY-1999;	9905-0134768;
PR	19-MAY-1999;	9905-0134591;
PR	20-MAY-1999;	9905-0135124;
PR	21-MAY-1999;	9905-0135353;
PR	24-MAY-1999;	9905-0135369;
PR	25-MAY-1999;	9905-0136021;
PR	27-MAY-1999;	9905-0136392;
PR	28-MAY-1999;	9905-0136782;
PR	01-JUN-1999;	9905-0137222;
PR	03-JUN-1999;	9905-0137528;
PR	04-JUN-1999;	9905-0137502;
PR	07-JUN-1999;	9905-0137724;
PR	08-JUN-1999;	9905-0138094;
PR	10-JUN-1999;	9905-0138540;
PR	10-JUN-1999;	9905-0138867;
PR	14-JUN-1999;	9905-0138119;
PR	16-JUN-1999;	9905-0138452;
PR	16-JUN-1999;	9905-0138453;
PR	17-JUN-1999;	9905-0138452;
PR	18-JUN-1999;	9905-0138454;
PR	18-JUN-1999;	9905-0138455;
PR	18-JUN-1999;	9905-0138461;
PR	18-JUN-1999;	9905-0138462;
PR	18-JUN-1999;	9905-0138463;
PR	18-JUN-1999;	9905-0138457;
PR	18-JUN-1999;	9905-0138750;
PR	18-JUN-1999;	9905-0139763;
PR	21-JUN-1999;	9905-0133887;
PR	22-JUN-1999;	9905-0133899;
PR	23-JUN-1999;	9905-0140353;
PR	23-JUN-1999;	9905-0140354;
PR	24-JUN-1999;	9905-0140695;
PR	28-JUN-1999;	9905-0140823;
PR	30-JUN-1999;	9905-0140991;
PR	01-JUL-1999;	9905-0141287;
PR	01-JUL-1999;	9905-0141842;
PR	01-JUL-1999;	9905-0142154;
PR	02-JUL-1999;	9905-0142055;
PR	06-JUL-1999;	9905-0142390;
PR	08-JUL-1999;	9905-0142803;
PR	09-JUL-1999;	9905-0142920;
PR	12-JUL-1999;	9905-0142977;
PR	13-JUL-1999;	9905-0143542;
PR	14-JUL-1999;	9905-0143624;
PR	15-JUL-1999;	9905-0144085;
PR	16-JUL-1999;	9905-0144086;
PR	19-JUL-1999;	9905-0144331;
PR	19-JUL-1999;	9905-0144332;
PR	19-JUL-1999;	9905-0144333;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145088;

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158993.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.98; Score 97.5; DB 21; Length 496;
Best Local Similarity 22.14; Pred. No. 0.027;
Matches 44; Conservative 30; Mismatches 86; Indels 39; Gaps 7;

OY 70 ISFDALFLGLIIRGYAALPNMQFSQFVKD-----LTKCYAMLLTYCGFVLDLVL 122
DB 312 ISFDALFLGLIIRGYAALPNMQFSQFVKD-----LTKCYAMLLTYCGFVLDLVL 365
OY 123 YHLNPKRSILEDYHEFLNINISFRNFIFAPITEEIFYTSMLTYLNL- PHSQLSYOO 181
DB 366 -----gsLftemyfvtfswnykyvygfmllvftvllvltvcvtlvgtyfllnaen 418
OY 182 LFQOPSLFFGLAHNAHAYEOLQEGSMVTYSILTTQCFQILYTTLEGG- TKFVFEVTRGG 239
DB 419 yhwqtsffsaa-----stavyvylsylyyyvktkmsgffqtsfyf---gy 462
OY 240 NLWCIIILHLCNIMGFRG 258
DB 463 tmmfcLgIlgIcgavgyIg 481

Search completed: February 12, 2002, 09:27:26
Job time: 112 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:32:01 ; Search time 43.94 Seconds

(without alignments)
1048.607 Million cell updates/sec

Title: US-09-165-460a-4

Perfect score: 1651
Sequence: 1 MGFSTFLVLYISISVLP.....ISLKDPLQTLVGTGPGYRITLL 315

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227.5	13.8	314	11 Q9CSF8	Q9CSF8 mus musculus
2	207	12.5	225	11 Q9XK03	Q9XK03 mus musculus
3	113.5	6.9	339	5 Q18364	Q18364 caenorhabditis
4	111.5	6.8	225	8 Q9XNM8	Q9XNM8 ceratitidis
5	109	6.6	1363	5 Q10031	Q10031 caenorhabditis
6	104.5	6.3	255	5 Q9PKE0	Q9PKE0 chlamydia m
7	102.5	6.2	137	5 Q9NM17	Q9NM17 leishmania
8	102.5	6.2	1471	3 Q74805	Q74805 schistosom
9	99.5	6.0	224	8 Q9XMI3	Q9XMI3 penaeus not
10	98	5.9	256	2 Q84256	Q84256 chlamydia t
11	97.5	5.9	289	5 O17359	O17359 caenorhabditis
12	97.5	5.9	315	11 Q9EPB8	Q9EPB8 mus musculus
13	97.5	5.9	351	5 Q16945	Q16945 aplysia cal
14	97.5	5.9	433	12 Q84483	Q84483 paramyosin
15	97.5	5.9	589	10 Q04091	Q04091 arabidopsis
16	97	5.9	483	2 Q9PPB9	Q9PPB9 campylobact
17	96.5	5.8	224	8 Q9MGN0	Q9MGN0 drosophila
18	96	5.8	328	2 Q9CEY3	Q9CEY3 lactococcus
19	96	5.8	333	4 Q9H2L2	Q9H2L2 homo sapien

20	96	5.8	346	8 Q9TCL5	Q9TCL5 gymnostinop
21	96	5.8	346	8 Q9G8K3	Q9G8K3 psarocollus
22	95.5	5.8	224	8 Q9MD63	Q9MD63 drosophila
23	95.5	5.8	434	11 Q54844	Q54844 mus musculus
24	95	5.8	274	2 Q9KFR4	Q9KFR4 bacillus ha
25	95	5.8	358	8 Q9GHR5	Q9GHR5 penum bold
26	95	5.7	500	5 Q22904	Q22904 caenorhabditis
27	94.5	5.7	224	8 Q9MGL1	Q9MGL1 drosophila
28	94.5	5.7	244	2 Q05525	Q05525 bacillus su
29	94.5	5.7	269	5 Q62262	Q62262 caenorhabditis
30	94.5	5.7	386	3 Q93875	Q93875 candida alb
31	94	5.7	346	5 Q16932	Q16932 caenorhabditis
32	94	5.7	347	2 Q9ZDJ0	Q9ZDJ0 rickettsia
33	94	5.7	432	10 Q24406	Q24406 arabidopsis
34	94	5.7	554	5 Q9BPP5	Q9BPP5 bombyx mand
35	94	5.7	769	3 Q12244	Q12244 saccharomyc
36	93.5	5.7	224	8 Q9MCK7	Q9MCK7 drosophila
37	93	5.6	645	8 Q63620	Q63620 balanogloss
38	92.5	5.6	224	8 Q9MJC8	Q9MJC8 drosophila
39	92.5	5.6	224	8 Q9MGN5	Q9MGN5 drosophila
40	92.5	5.6	411	4 Q9Y2E8	Q9Y2E8 homo sapien
41	92.5	5.6	437	4 Q9Y507	Q9Y507 homo sapien
42	92.5	5.6	670	8 Q9T447	Q9T447 dictyostell
43	91.5	5.5	449	4 Q43934	Q43934 homo sapien
44	91.5	5.5	449	4 Q9NX15	Q9NX15 homo sapien
45	91.5	5.5	457	8 Q47876	Q47876 alligator m

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	314 AA.
Q9CSF8	Q9CSF8			
AC	Q9CSF8			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	RAS AND A-FACTOR-CONVERTING ENZYME 1 HOMOLOG (S. CEREVISIAE)			
DE	(FRAGMENT).			
GN	RCE1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SYRAIN-C57BL/6J; TISSUE=EMBRYO;			
RX	MEDLINE=21085660; Pubmed=11217851;			
RA	Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,			
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Knehl P., Lewis S., Matsuo Y., Nikaido I., Pospole G., Quackenbush J.,			
RA	Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,			
RA	Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldi M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.F.,			
RA	Guinacchini S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mezzarrelli J., Mombaerts P.,			
RA	Mordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL: AK012946; BAB28566.1; -			
DR	MED: MG1:1336895; Rcel.			
FT	NON_TER			
SO	SEQUENCE	314 AA;	34390 MW;	DE2624BB17E234B5 CRC64;

```

Query Match      13.8%; Score 227.5; DB 11; Length 314;
Best Local Similarity 30.2%; Pred. No. 1.6e-11;
Matches 84; Conservative 41; Mismatches 82; Indels 71; Gaps 16;

OY 4 FSTFLVLLYISIVLPYATSOPEGRKDNPRITKSMOKLTIMLSNLF-LVPELOSQ 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 FSCF-----SLACSYVGSILYVKSE--LPKDHPAVYIKRR--STSVLVSSLSPLCVLWRE 76

OY 63 LSSFT---TSHISFMDAPLGLGIIIPGYAALPNPQSFQVVDLTKCYAMLT--LYCGPV 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 LTGIQPGTSLTLTM-GRLEGIIPA--ALLP-----LLTMTILFLGPL 116

OY 118 LDFV-----LYHLNPKSSILEDYHEFLINMS-----FRNFIFAPITEIEF 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 MQLSMDCPDLTDGLKAVLARS-----NARCLTDMRWLRNNOYIAPITELEV 163

OY 160 YTSMLTLYLNLIPHSOLSYQQLFWQPSLFGFLAHNAHAYEOLQSGMTTVSILLTTCFQ 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 FRACMLPR---LAPCTGLG-PAVFTCP-LFEGVAHFHIIIEQLRPROSSVGSIFVSAFQ 218

OY 220 ILTYTLTGGLTKFVFRVGGNLMCCIIILHALCINMGFP 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 FSYTAVFGAYTAFLEFIR-GLHIGPVLCHSFCNYMGFP 235

RESULT 2
O99K03 PRELIMINARY; PRT; 225 AA.
AC O99K03;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:8174).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004060; AAH04060.1;
SQ SEQUENCE 225 AA; 24791 MW; EFDDC2B6CE205B45 CRC64;

Query Match      12.5%; Score 207; DB 11; Length 225;
Best Local Similarity 31.8%; Pred. No. 6e-10;
Matches 63; Conservative 27; Mismatches 50; Indels 58; Gaps 10;

OY 80 GIIPGYAALPNPQSFQVVDLTKCYAMLT--LYCGPVLDV-----LYHLN 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 GIIPA--ALP-----LLTMTILFLGRLMQLSMDCPDLTDGLKAVLA 47

OY 128 PKSSILEDYHEFLINMS-----FRNFIFAPITEIEFTSMILTLYLNLIPHSOLSY 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 PRS-----NARCLTDMRWLRNNOYIAPITELEVRAACMLPR---LAPCTGLG- 90

OY 180 QQLFWQPSLFGFLAHNAHAYEOLQSGMTTVSILLTTCFQILLTYTLTGGLTKFVFRVGTG 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 PAVFTCP-LFEGVAHFHIIIEQLRPROSSVGSIFVSAFQSYTAVFGAYTAFLEFIR-G 148

OY 240 NLMCCIIILHALCINMGFP 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 HLIGPVLCHSFCNYMGFP 166

RESULT 3
O18364 PRELIMINARY; PRT; 339 AA.
AC O18364;
DT 01-NOV-1996 (Tremblrel. 01, Created)

```

```

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DB C33A12.10 PROTEIN.
CN C33A12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; Pubmed=7906398;
RP SEQUENCE FROM N.A.
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Paterson J., Percy C., Rifken L., Koopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
DR EMBL; Z68493; CAA92793.1;
DR InterPro; IPR003839; DUF215.
DR Pfam; PF02688; DUF215; 1.
SQ SEQUENCE 339 AA; 39177 MW; 05B7B6472ADC7DD5 CRC64;

Query Match      6.9%; Score 113.5; DB 5; Length 339;
Best Local Similarity 22.3%; Pred. No. 0.07;
Matches 70; Conservative 41; Mismatches 102; Indels 101; Gaps 17;

OY 3 QFSFVLV-----YISIVLPYATRS--QPEGRKDNPRITKSMOKLTIMLS 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 QFSFLAIVPWIVIPSEFVICYLFKILYTSNMKSEGRKQHVFLVIS-LSQFTCL- 84

OY 52 NLFVLPFLOSLSSTSHISF-----KDAFLGLGIIIPGY--YALPNPW----- 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 -LEFLDLMPLRPSTGTGFTSCASIPRNHYIKIIVSLAHTNYLAMSFPFLPIVRLIIV 143

OY 94 -----QFSQFVKDLTKCYAMLTLYCGVLDVLYHLNPKSSILEDYHEFLN 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 ICPKSHSKININIVKFAIPILCPFLCTEYFIPALG-VCKQFATP-----YFGS 193

OY 143 IWSFRNFIFAPITEIEFTSM--LTTLYLNLIPHSOLSYQQLFWQPS-----LEFGL 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 VW-----IYIISAFGLRNSFPL-----YNIIVPMIASITANGLEFFKV 233

OY 193 AHAHAYEOLQSGS--MTTVSILLTTCFQIL-YTTLFGGLTKFV-----VR 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 AQAARAKLISAGSGSKTHISITMTLAMIIFLTNCAFLLVYIFSGTSYISYSELIR 293

OY 237 TGNL-LWCCIIILHA 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 SFGNDLQFCVYVWA 307

RESULT 4
O9XN8 PRELIMINARY; PRT; 225 AA.
AC O9XN8;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34).
GN ATPASE 6.
OS Ceratitis capitata (Mediterranean fruit fly).

```


OG Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritoidea; Tephritidae; Ceratitis.
 NCBI_Taxid=7213;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP Spanos L., Koutroubas G., Kotsyakis M., Louis C.;
 RT "The mitochondrial genome of the Mediterranean fruitfly, *Ceratitis capitata*,"
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL, IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS, ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE ATPASE A CHAIN FAMILY.
 DR EMBL: AJ242872; CAB45092.1; -
 DR InterPro: IPR000568; ATP_synth_A.
 DR Pfam: PF00119; ATP_synth_A.1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASE_A.1.
 KM CF(0); Hydrogen ion transport; Hydrolase; Mitochondrion;
 KM Transmembrane.
 SQ SEQUENCE 225 AA; 25113 MW; BC6D91F3AF3A915 CRC64;

Query Match 6.8%; Score 111.5; DB 8; Length 225;
 Best Local Similarity 22.0%; Pred. No. 0.067;
 Matches 58; Conservative 40; Mismatches 97; Indels 69; Gaps 12;

QY 49 LISULFLVPLFQISQSLSTSHISFKDAFLGIGIPGYALPNPWOEQYVKDLTKCYAM 108
 DB 1 MATNLFSVFDPSSSIFNIS--LNMSTFLGILLIPSAWMLPSPW----- 43
 QY 109 LLLTYCGVLDVLYHLNPKSSILEDYHELMN-----WSFNFIAPITTEIFYSM 163
 DB 44 -----HIF--WNSILMTLHKEFKTLGSGHSGSTFEIVSL-----FSLI 81
 QY 164 LLLTYLNLIPHSQSLYQOL-----FMQPSLFFG-LAHANHAVEQLQSGSMITVSIL 213
 DB 82 LFNFMGLFPIYITSTSLTLTLALPLWLCFMYLGYINFTQMFHTLVYQGTIPAVLMP 141
 QY 214 LTTGQIILYTLFGGLKRFVVRPGNLMCCIIILHALCINMGFPGPSRLNLHFTVVD-KK 272
 DB 142 FMVCIETISNIRPGT--LAVRLAAM--IAGHLTLTLTGNTGPS--LSFMIVSILL 192
 QY 273 AGRISKL-----VSIMNKCYSALL 291
 DB 193 IGOIALLVESAVALIQSYFAVL 216
 RESULT 5
 ID 010031 PRELIMINARY; PRT: 1363 AA.
 AC Q10031;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE C27D6.2 IN CHROMOSOME II
 DE (NC 2.7.1.-).
 GN C27D6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_Taxid=6239;
 RN [1]
 RA SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Ding H.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: U23179; AAC46722.1; -
 DR WormPep: C27D6.2; CE01812.
 DR InterPro: IPR002106; AA_tRNA_ligase-II.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002290; Ser_thr_kin_acctsite.
 DR InterPro: IPR002184; Srb.
 DR Pfam: PF00069; Kinase.1.
 DR Pfam: PF02175; Srb.5.
 DR SMART: SM00220; S_TKC.1.
 DR PROSITE: PS00339; AA_tRNA_LIGASE-II.2; UNKNOWN.4.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN.1.
 DR PROSITE: PSS0011; PROTEIN KINASE DOM.1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
 KM Transferase; Transmembrane.
 FT TRANSMEM 22 42 POTENTIAL.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.
 FT TRANSMEM 380 400 POTENTIAL.
 FT TRANSMEM 415 435 POTENTIAL.
 FT TRANSMEM 460 480 POTENTIAL.
 FT TRANSMEM 498 518 POTENTIAL.
 FT TRANSMEM 544 564 POTENTIAL.
 FT TRANSMEM 592 612 POTENTIAL.
 FT TRANSMEM 635 655 POTENTIAL.
 FT TRANSMEM 671 691 POTENTIAL.
 FT DOMAIN 679 682 POLY-LEU.
 FT TRANSMEM 716 736 POTENTIAL.
 FT TRANSMEM 734 774 POTENTIAL.
 FT TRANSMEM 832 852 POTENTIAL.
 FT TRANSMEM 867 887 POTENTIAL.
 FT TRANSMEM 913 933 POTENTIAL.
 FT TRANSMEM 952 972 POTENTIAL.
 FT TRANSMEM 999 1019 POTENTIAL.
 FT DOMAIN 1068 1363 POTENTIAL.
 FT NP_BIND 1074 1082 ATP (BY SIMILARITY).
 FT ACT_SITE 1193 1193 BY SIMILARITY.
 FT BINDING 1097 1097 ATP (BY SIMILARITY).
 SQ SEQUENCE 1363 AA; 157850 MW; 7A030E80C96CC0A6 CRC64;

Query Match 6.6%; Score 109; DB 5; Length 1363;
 Best Local Similarity 21.5%; Pred. No. 0.73;
 Matches 67; Conservative 46; Mismatches 110; Indels 88; Gaps 16;

QY 18 VLPIKATSPGSKRDNPRITKSMOKITIMLISNLPVLPLOSLSSTSHISFKDAFL 77
 DB 219 VAPLEITSMCD-----LITRSLYK-----VGNLSLTLMTQM----- 252
 QY 78 GLGIIPGYAALPNPWOEQYVK-DLTKCYAMLLTYCGVLDV-----VLHYLNP 128
 DB 253 -----IMPLGFSIERIALSMKTSYENVRT-FLGPILVYTLIGIDALLHYVFRD 301
 QY 129 KSSILEDYHEHELNWSEFRNFIAPITE-----EIFYSM-LTTYLNLIPHSQ 177
 DB 302 EK-FED-----SPISFALVPERSAIPFNSYFELLVAEIGNFCIFILVHSKF 350
 QY 178 S----YQQLFWOPS---LFGLAHHAHVEQLQSGSMITVSILLTTCQIILYTLFGGLTK 231
 DB 351 KARVHOMLENDVSCVCELAQOLAY-HPVYRSSQFMSWLVSSISIPALYFIRKIF----- 404
 QY 232 FVFPRTGNLMCCIIILHALCNI-----MGFPGPSRLNLHFTVVDKKGRISKLISYNNK- 285
 DB 405 --FLHFGNKLCLLIYFICHLFSLMALCFAPFYQFLIPFVTSKQOLLNITTLFKNGQI 462
 QY 286 CYFALLVGLI 296
 DB 463 CSFLLITSSML 473


```

Db 579 FISLONSKNEFIYLLDFY--MOLVSSONKNPKRTAISRRKPTLSMKSSSL--SOLMT 633
OY 62 OLSSTSHISKDAFLGILIPGYAAL-PN-----PMQSFQVADLTFCVAMLLTY--C 114
Db 634 TVSSSTNVA-----YIRCIKPNKEKLPMTESP-----DPAVLSQRLAC 669
OY 115 GPVDFVLYHLNPKSSILEDYHEFLINWSPFNIFAPITEEIFYTSMLTTLNLIPLH 174
Db 670 GVFETIRSSIGFAPRSEYEEFAHRFRLLSSKEM-----EDDNKKTLTINVSATPH 722
OY 175 SOLSTQ-----OLFQPSLEFGLAHNAH-----YEQLOEGSMITVSIILTTTCFOI-L 221
Db 723 DNLNFQVRSKIFFRSNIIGNFEBAHRATCSKSTVLQSAIRGFEFTREYQRTVAFIKL 782
OY 222 YTLTFCGLTKFVFVR 236
Db 783 QSVIMGWLTTRKPER 797

RESULT 9
O9XMT3 PRELIMINARY; PRT; 224 AA.
AC O9XMT3;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34).
GN ATP6.
OS Penaeus notialis.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC Penaeidae; Penaeus.
OX NCBI_TaxID=44108;
RN [1]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E., Denneboudy N., Oliva-Suarez M., Mounolou J.C.,
RA Monnerot M.;
RT "Partial sequence of the shrimp Penaeus notialis mitochondrial
RT genome.";
RL C. R. Acad. Sci., D, Sci. Nat. 319:473-486(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E.;
RL Thesis (1997), Paris-Sud, UFR d'Orsay, France.
RN [3]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E., Pempera M., Denneboudy N., Oliva-Suarez M.,
RA Mounolou J.C., Monnerot M.;
RT "Mitochondrial genes collectively suggest the paraphyly of Crustacea
RT with respect to Insecta.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE
CC (BY SIMILARITY).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS. CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ATPASE A CHAIN FAMILY.
DR EMBL: X84350; CAB40367.1; -.
DR InterPro: IPR000568; ATP_synth_A.
DR Pfam: PF00119; ATP_synth_A; 1.
DR PRINTS: PR00123; ATPASEA.
DR PROSITE: PS00449; ATPASE_A; 1.
KM CF(0): Hydrogen ion transport; Mitochondrion; Transmembrane.
SQ SEQUENCE 224 AA; 24744 MW; E903131029A715A5 CRC64;

```

```

OY 51 SNLFVPELQSLSTSHISKDAFLGILIPGYAALPNPWO--FSQFVKDLTKCYAM 108
Db 13 SSTMLP-----LWVSTFLGVMLPLMYAMPSPBSMLMLVATLHKERT 60
OY 109 LL-FLYCGPVLDV-LVHLNPKSSILEDYHEFLINWSPFNIFAPITEEIFYTSMLLT 166
Db 61 LIGSSHFGTTLMTFLSFLSI-----VFNNFLGLLP--YIFTS-----TSHLVW 101
OY 167 TYLNLIPHSQSLSYQOLFQPSLFFG-LAHNAHAYEQ-----QEGSMITVSIILTTTCFOI-L 221
Db 102 TSLALP-----LWVAFMLFGWINHTOHMFANLTPGSTGALMPFVLIETISNVI 152
OY 222 YTLTFCGLTKFVFRTGCLMCCITLHALCNMGPGSRILNHTVYDKKAGRSKLVS 281
Db 153 RP--GTLA--VRLAANM--IAGHLLTLTGSTGS--LSATLI----- 187
OY 282 IWNKCYFALLVGLISL 298
Db 188 -----SMLITGQILL 197

RESULT 10
O84256 PRELIMINARY; PRT; 256 AA.
AC O84256;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHETICAL 28.4 KDA PROTEIN.
GN CT254.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; Pubmed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AE001298; AAC67847.1; -.
DR InterPro: IPR003675; AD1.
DR Pfam: PF02517; AD1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 256 AA; 28378 MW; 25BBE45DE4B53B83 CRC64;

```

Query Match 6.0%; Score 99.5; DB 8; Length 224;
 Best Local Similarity 22.2%; Pred. No. 0.68;
 Matches 57; Conservative 41; Mismatches 78; Indels 81; Gaps 15;

Query Match 5.9%; Score 98; DB 2; Length 256;
 Best Local Similarity 22.9%; Pred. No. 1;
 Matches 60; Conservative 34; Mismatches 110; Indels 58; Gaps 9;

Db 228 FIVEKVRHIAAPVTLHLFNSC 249

RESULT 11

017359 PRELIMINARY; PRT; 289 AA.
 AC 017359;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE COSE4.5 PROTEIN.
 GN COSE4.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latierille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smidson N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RN Nature 368:32-38(1994).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Blanchard M., Kramer J., Gibson A.,
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.,
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF026209; AAB71274.1;
 DR InterPro: IPR000168; 7TM_nematode.
 DR InterPro: IPR003002; 7TM_chemorecept_1.
 DR Pfam: PF01461; 7tm_4; 1.
 SO SEQUENCE 289 AA; 32908 MW; FFC7953558200B31 CRC64;

Query Match 5.9%; Score 97.5; DB 5; Length 289;
 Best Local Similarity 23.4%; Pred. No. 1.3;
 Matches 60; Conservative 26; Mismatches 71; Indels 99; Gaps 13;

QY 95 FSQEVKDLTKCVAMLLTGYCPVLDVLYHLINPKSSILEDYH---EFLINMSFRNFI 150
 DB 15 FSQFIVY-----FILLYLAKAG-----KHLGPIGFKNLSEF-NFF 49
 QY 151 FAPITEEFY-----TSMILT-----TYNLIRPHQSLSYQOLF----- 183
 DB 50 Y-----TFYLGIFGASGLCTSLCTEFMRFRYIYCOOKYLLHLLEGRKLLLFLEPTTI 103
 QY 184 ---NOPSIFGLAHNAHAYEOL---QEGSMWTVSILLTFOILYTLTLEGLTFKVFVR 236
 DB 104 SITWTCGFCGLITSEKQELRIPQENFEGDSNVLMFVSGY-----WIRGA 152
 QY 237 TGNLMC-----CIILALCNMGFPGPSRLNLHFTVVDKAGRISKLVSIMNKCYPAL 290
 DB 153 NGEKIMCLRDGFGFLGICGCSVISFCGLKT---FKKMIIEVQSGMSQTAALNKLQPLT 208
 QY 291 LVLGISLAKOTLQTLV 306
 DB 209 L-----TLQTL 215

RESULT 12

09EPB8 PRELIMINARY; PRT; 315 AA.
 AC 09EPB8;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE VN7 (VOMERONASAL RECEPTOR VIR9).
 GN VIR9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lane R.P., Rowen L., Friedman C., Trask B.J., Hood L.,
 RT "Genomic characterization of the murine chromosome 6D1 VNO olfactory
 RT receptor cluster."
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SVJ;
 RX MEDLINE=20568485; PubMed=11116090;
 RA Del Punta K., Rothman A., Rodriguez I., Mombaerts P.,
 RT "Genes in the Mouse."
 RL Genome Res. 10:1958-1967(2000).
 DR EMBL: AF129005; AAG3282.1;
 DR EMBL: AF291488; AAG42082.1;
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR PROSITE: PS50262; G_PROTEIN_RECEPT_FL2; 1.
 KW Receptor.
 SO SEQUENCE 315 AA; 35853 MW; 1D6A41D0FEBB6C93 CRC64;

Query Match 5.9%; Score 97.5; DB 11; Length 315;
 Best Local Similarity 20.4%; Pred. No. 1.4;
 Matches 69; Conservative 50; Mismatches 111; Indels 109; Gaps 15;

QY 5 STEPLVLYSISVPLVATYQPEGSKRDNPRTKSRMOKLTMLISNLVLPPLQSL 64
 DB 28 NSFLFLFHI-FKYI-----RGQRSHRIDPILGLSLIHLVLMILAMSLV 69
 QY 65 STTHSHSKDAFLGILGILGYAALPNW-----QFQFVADLTKCVAMLLTL 112
 DB 70 AT-----DIFW-----FWGRMGDTCKCVISLYRPFCSLSLCATSLTS- 107
 QY 113 YCGPVLDFVLYHLINPKSSILEDYH-----FLINMSFRNFIAPITEIEFYS 162
 DB 108 ----ILQAV---TLNPRNSCLEKFRKSPHYMGLCLFLSV--FYTFSSPLATYITKS 158
 QY 163 ML-----LTYNLIRPHQSLSYQOLF---WOPSIFGLAHNAHAY-----EQL 202
 DB 159 NLTPSPFYITTSCLAPMSYFHLTWFILITSRDVIFVGLMLSSGYMTFLGRHKKS 218
 QY 203 QEGSMWTVS-----ILLTFOILYTLTLEGLTFKVFVRFGMLMCCIL- 247
 DB 219 QPLHTSFLKPSAEKRAMRTILCLMSFFVLMY--LDSIVSYIRSDIDGQIFCVHIFT 276
 QY 248 -HALCNMGFPGPSRLNLHFTVVDKAGRISKLVSIMNKCYPAL 285
 DB 277 AHGYATVSPFLITSTEKYIINFRSTGRMVTITILLRR 315

RESULT 13
 Q16945 PRELIMINARY; PRT; 351 AA.
 AC 016945;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

Query Match	5.9%;	Score 97.5;	DB 5;	Length 351;
Best Local Similarity	25.4%;	Pred. No. 1.6;		
Matches 61; Conservative	30;	Mismatches 84;	Indels 65;	Gaps 13

```

query match      5.98;  Score 97.5;  DB 12;  Length 433;

```

RESULT	15	
ID	004091	
PRELIMINARY:		
PRT:	589	AA.

Query Match	5.9%	Score	97.5	DB	10	Length	589
Best Local Similarity	22.1%	Pred. No.	2.8				
Matches	44	Conservative	30	Mismatches	86	Indels	39
						Gaps	7

[illegible]

Tue Feb 12 16:22:39 2002

us-09-165-460a-4.rpt

Page 8

Search completed: February 12, 2002, 09:32:03
Job time: 350 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:07 ; Search time 21.66 Seconds
(without alignments)
327.264 Million cell updates/sec

Title: US-09-165-460a-4

Perfect score: 1651
Sequence: 1 MLOFSTFVLVLYISISYLP.....ISLKDTLQTLVGNPGYRITL 315

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 21252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	228.5	13.8	329	US-09-022-669-2	Sequence 2, Appli
2	227.5	13.8	329	US-09-262-749-2	Sequence 2, Appli
3	96.5	5.8	456	US-09-058-389A-2	Sequence 2, Appli
4	91.5	5.5	470	US-09-292-071-25	Sequence 25, Appli
5	91.5	5.5	470	US-09-292-069A-25	Sequence 25, Appli
6	91.5	5.5	471	US-07-817-920-8	Sequence 8, Appli
7	91.5	5.5	471	US-08-370-542-7	Sequence 8, Appli
8	91.5	5.5	471	US-08-117-006-8	Sequence 8, Appli
9	91.5	5.5	471	US-08-216-594-8	Sequence 8, Appli
10	91.5	5.5	471	US-08-542-358-7	Sequence 7, Appli
11	91.5	5.5	471	US-08-244-434-2	Sequence 2, Appli
12	91.5	5.5	471	US-09-018-351-7	Sequence 7, Appli
13	91.5	5.5	471	US-09-032-742-8	Sequence 8, Appli
14	91.5	5.5	471	PCT-US93-00149-8	Sequence 8, Appli
15	91.5	5.5	471	US-07-996-772A-11	Sequence 11, Appli
16	89.5	5.4	471	US-09-032-742-2	Sequence 2, Appli
17	89	5.4	3174	US-08-477-451-3	Sequence 3, Appli
18	87.5	5.3	445	US-08-630-118A-2	Sequence 2, Appli
19	87.5	5.3	445	US-08-838-399-2	Sequence 2, Appli
20	87.5	5.3	445	US-09-235-839-2	Sequence 2, Appli
21	84	5.1	462	US-08-477-451-24	Sequence 2, Appli
22	83.5	5.1	445	US-08-630-118A-4	Sequence 4, Appli
23	83.5	5.1	445	US-08-630-118A-6	Sequence 6, Appli
24	83.5	5.1	445	US-08-838-399-4	Sequence 4, Appli
25	83.5	5.1	445	US-08-838-399-6	Sequence 6, Appli
26	83.5	5.1	445	US-09-003-199-2	Sequence 2, Appli
27	83.5	5.1	445	US-09-003-199-21	Sequence 21, Appli

28	83.5	5.1	445	US-09-003-199-23	Sequence 23, Appli
29	83.5	5.1	445	US-09-235-839-4	Sequence 4, Appli
30	83.5	5.1	445	US-09-235-839-6	Sequence 6, Appli
31	83.5	5.1	445	US-09-040-958-2	Sequence 2, Appli
32	83.5	5.1	445	US-09-040-958-4	Sequence 4, Appli
33	83.5	5.1	445	US-08-349-025-4	Sequence 4, Appli
34	83.5	5.1	445	US-08-566-096A-4	Sequence 4, Appli
35	83.5	5.1	445	US-08-668-650B-4	Sequence 4, Appli
36	83.5	5.1	445	US-08-349-025-2	Sequence 2, Appli
37	83.5	5.1	445	US-08-566-096A-2	Sequence 2, Appli
38	83.5	5.1	445	US-08-668-650B-2	Sequence 2, Appli
39	83.5	5.1	445	PCT-US95-15646-2	Sequence 2, Appli
40	83.5	5.1	445	PCT-US95-15646-4	Sequence 4, Appli
41	83	5.0	821	US-09-472-869-24	Sequence 24, Appli
42	82.5	5.0	314	US-08-988-876-7	Sequence 7, Appli
43	82	5.0	379	US-08-118-270-32	Sequence 32, Appli
44	82	5.0	379	PCT-US93-08528-32	Sequence 32, Appli
45	81.5	4.9	326	US-09-058-389A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-022-669-2
Sequence 2, Application US/09022669
Patent No. 610717
GENERAL INFORMATION:
APPLICANT: KIRKLY, KRISTINE
APPLICANT: SOUTHAN, CHRISTOPHER
APPLICANT: KNAB, ANNE
TITLE OF INVENTION: Human RCE1
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.669
FILING DATE: 12-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 97304437.3
FILING DATE: JUNE 24, 1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70379
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-022-669-2
Query Match 13.8%; Score 228.5; DB 3; Length 329;
Best Local Similarity 30.6%; Pred. No. 6,4e-17;
Matches 83; Conservative 39; Mismatches 82; Indels 67; Gaps 15;

QY II LYSISVLP LYATSQPEGSK

7-4660C-9C0-60-50

OY |::||| ||||:| |:
I1 LYSISVLELYATSOPEGSRDNPRIKSRMOKLTIMLSNLF-LVPFLOSQSSSTTSH 69

RESULT 4
US-09-292-071-25

Sequence 25, Application us/09292071
Patent No. 6107324
GENERAL INFORMATION:
APPLICANT: Behan, Dominic
APPLICANT: Chalmers, Derick
TITLE OF INVENTION: No. 6107324-Endogenous, Constitutively Activated
TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators Thereof
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arena Pharmaceuticals, Inc.
STREET: 6166 Nancy Ridge Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/292,071
FILING DATE: April 14, 1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Rosen
REGISTRATION NUMBER: 39,822
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 564-6525
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-292-071-25

Query Match	5.5%;	Score 91.5;	DB 3;	length 470;
Best Local Similarity	20.4%;	Pred. No. 0.1;		
Matches	69;	Conservative	52;	Mismatches 77;
				Indels 141;
				Gaps 21

RESULT

```

US-09-292-069A-25
Sequence 25, Application US/09292069A
Patent No. 6140509

GENERAL INFORMATION:
APPLICANT: Behan, Dominic P
APPLICANT: Chalmers, Derek T
APPLICANT: Foster, Richard J
APPLICANT: Glen, Robert C
APPLICANT: Lawless, Michael S
APPLICANT: Liaw, Chen W
APPLICANT: Liu, Qian
APPLICANT: Russo, Joseph F
APPLICANT: Smith, Julian R
APPLICANT: Thomsen, William J
TITLE OF INVENTION: No. 6140509-Endogenous, Constitutively Activated Human
TITLE OF INVENTION: Serotonin Receptors And Small Molecule Modulators
TITLE OF INVENTION: thereof
FILE REFERENCE: AREN0033
CURRENT APPLICATION NUMBER: US/09/292,069A
CURRENT FILING DATE: 1999-04-14
PRIOR APPLICATION NUMBER: 60/090,783
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/112,909
PRIOR FILING DATE: 1998-12-18
PRIOR APPLICATION NUMBER: 60/123,000
PRIOR FILING DATE: 1999-03-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 470
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: No. 6140509seq
US-09-292-069A-25

```

Query Match	5.5%;	Score 91.5;	DB 4;	Length 470;
Best Local Similarity	20.4%;	Pred. No. 0.1;		
Matches 69;	Conservative 52;	Mismatches 77;	Indels 141;	Gaps 21;

QY	45	LTIMLINELFLVPLQSLSTTSH	-----ISFKATIGLGITP	-----Gy	85
		:	: : : :		
Db	93	LVTMAVS-----LEKKLQATNATFYMSLAIDMLGFLMPVSM	LTILGYRWPLPSK	145	
QY	86	-----YALPNPWPQSQ	-----VKDLTKCYAMLLTYCG	115	

Db 146 LCAWIVYLDVLFSTASIMHLCALISDRVYALONPIHHSRFSNRTKAFKIIA-VWTISVG 204
Qy 116 PVLDVLYHLNPK-----SILEDYHEFLNINSEFNFTFAPITEEFTSMLLTYL 169
Db 205 ISMPIPVGLDDSKVEKESGLAD--DNEVLIGSFVSF-FIPU-----TIMVITYFL 255
Qy 170 NL-----IPHSOLSYOOLFW-----QPSLFGLAHAHAY 199
Db 256 TIKSLQKATLCVSDLGTRAKLASFLPQSSLSSEKLFORSIHREPGSYTG----- 307
Qy 200 EQLQGSMTVSIILTTGFOILYTLTFGLTKFVFVRGTGNLMC-----CITLHALC 251
Db 308 ----RRTMOSISNEQKAC-KYL-----GIVFELFV-----VAMCFFITINIMAVICKESC 352
Qy 252 N--IMGFPGSRNLHFTVVDKKGRIKLV-SIWNKCY 287
Db 353 NEDVIG----ALLNV-FWIGYLSAVNPPLYTTFENKTY 386

RESULT 6

US-07-817-920-8
Sequence 8, Application US/07817920
Patent No. 5360735

GENERAL INFORMATION:

APPLICANT: Weinschank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
NUMBER OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York.
STATE: New York.
COUNTRY: USA
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/817,920
FILING DATE: 19920108
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39318
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: Linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
IMMEDIATE SOURCE:
CLONE: 5-HT2
US-07-817-920-8

Query Match 5.5%; Score 91.5; DB 1; Length 471;
Best Local Similarly 20.4%; Pred. No. 0.1;
Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

Qy 45 LTIMLSMLFLVPFQSSLSSTSH---ISFKQAFGLGIIP-----CY----- 85
Db 94 LVIMAVS-----LEKRLQATNTYFENLSIADLLGFLVPMYSMLTILGYRPLDSK 146
Qy 86 -----YALPNPQFQFQF---VKDLTKCVAMLLTYCG 115
Db 147 LCAWIVYLDVLFSTASIMHLCALISDRVYALONPIHHSRFSNRTKAFKIIA-VWTISVG 205
Qy 116 PVLDVLYHLNPK-----SILEDYHEFLNINSEFNFTFAPITEEFTSMLLTYL 169
Db 206 ISMPIPVGLDDSKVEKESGLAD--DNEVLIGSFVSF-FIPU-----TIMVITYFL 256
Qy 170 NL-----IPHSOLSYOOLFW-----QPSLFGLAHAHAY 199
Db 257 TIKSLQKATLCVSDLGTRAKLASFLPQSSLSSEKLFORSIHREPGSYTG----- 308
Qy 200 EQLQGSMTVSIILTTGFOILYTLTFGLTKFVFVRGTGNLMC-----CITLHALC 251
Db 309 ----RRTMOSISNEQKAC-KYL-----GIVFELFV-----VAMCFFITINIMAVICKESC 353
Qy 252 N--IMGFPGSRNLHFTVVDKKGRIKLV-SIWNKCY 287
Db 354 NEDVIG----ALLNV-FWIGYLSAVNPPLYTTFENKTY 387

RESULT 7

US-08-370-542-7
Sequence 7, Application US/08370542
Patent No. 5476782

GENERAL INFORMATION:

APPLICANT: Weinschank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND
NUMBER OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,542
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/194,113
FILING DATE:
APPLICATION NUMBER: US/07/803,626

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39318
TELECOMMUNICATION INFORMATION:

RESULT 11
 US-08-244-434-2
 Sequence 2, Application US/08244434
 Patent No. 5854004
 GENERAL INFORMATION:
 APPLICANT: Chernilofsky, A.P.
 APPLICANT: Himmel, A.
 APPLICANT: Stratowa, C.
 APPLICANT: Meyer, U.
 APPLICANT: Lamche, H.
 TITLE OF INVENTION: Process for Screening Substances Capable of
 Modulating a Receptor-Dependent Cellular Signal
 TITLE OF INVENTION: Transmission Path
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244,434
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/92/02718
FILING DATE: PCT Filing Date: 25-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1340000/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-244-434-2

Query Match 5.5%; Score 91.5; DB 2; Length 471;
Best Local Similarity 20.4%; Pred. No. 0.1;
Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

QY 45 LITMLINLFLVPLQSLSTSH---ISFKAFGLGIIP-----GY----- 85
DB 94 LVIMAVS-----LEKRLQATNYFLMSLAIDMLGFLVMPVSMULTIYGYWPLPSK 146
QY 86 -----YAAPNPWFQSF---VKDLTKCVAMLLTYCG 115
DB 147 LCAWITLDVLFSTASIMHLCATSLDRVAIONIHHSRFRSKRAKLKITA-VWITSVG 205
QY 116 PVLDVLYHLNPK-----SILEDIFYHEFLINWFRNFIFADITEIEFTYMLTTYL 169
DB 206 ISMIPVFGLODDSKVKEGSCLLAD--DNFVLIGSFVS-FIPL-----TIWVITYFL 256
QY 170 NL-----IPHSQSLYQQLFW-----QPSLFFGLAHAHAY 199
DB 257 TITSLOEARLQVSDICTRAKLASFSLPOSSLSSEKLFORSIHREGSYTG----- 308
QY 200 EQLOEGSMFTVSIILTTGFOILYTLTFLGGLTKFEVFRGGLNLC-----CIILHALC 251
DB 309 -----RRTMOSISNQKAC-KYL-----GIVFELFV---VMGCPFITINIMAVICKESC 353
QY 252 N-IMGFPGPSRLNLHFTVYDKKAGRISKLV-SIWNKCY 287
DB 354 NEDVIG---ALLNV-FVMIGYLISSAVNPVLYTLEFNKTY 387

RESULT 12
US-09-018-351-7
Sequence 7, Application US/09018351
Patent No. 6096507
GENERAL INFORMATION:
APPLICANT: Weinshank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,351
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/542,358
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317-22/JPM/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-09-018-351-7

Query Match 5.5%; Score 91.5; DB 3; Length 471;
Best Local Similarity 20.4%; Pred. No. 0.1;
Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

QY 45 LITMLINLFLVPLQSLSTSH---ISFKAFGLGIIP-----GY----- 85
DB 94 LVIMAVS-----LEKRLQATNYFLMSLAIDMLGFLVMPVSMULTIYGYWPLPSK 146
QY 86 -----YAAPNPWFQSF---VKDLTKCVAMLLTYCG 115
DB 147 LCAWITLDVLFSTASIMHLCATSLDRVAIONIHHSRFRSKRAKLKITA-VWITSVG 205
QY 116 PVLDVLYHLNPK-----SILEDIFYHEFLINWFRNFIFADITEIEFTYMLTTYL 169
DB 206 ISMIPVFGLODDSKVKEGSCLLAD--DNFVLIGSFVS-FIPL-----TIWVITYFL 256
QY 170 NL-----IPHSQSLYQQLFW-----QPSLFFGLAHAHAY 199
DB 257 TITSLOEARLQVSDICTRAKLASFSLPOSSLSSEKLFORSIHREGSYTG----- 308
QY 200 EQLOEGSMFTVSIILTTGFOILYTLTFLGGLTKFEVFRGGLNLC-----CIILHALC 251
DB 309 -----RRTMOSISNQKAC-KYL-----GIVFELFV---VMGCPFITINIMAVICKESC 353
QY 252 N-IMGFPGPSRLNLHFTVYDKKAGRISKLV-SIWNKCY 287
DB 354 NEDVIG---ALLNV-FVMIGYLISSAVNPVLYTLEFNKTY 387

RESULT 13
US-09-032-742-8
Sequence 8, Application US/09032742
Patent No. 6255089
GENERAL INFORMATION:
APPLICANT: Teitel, Milt
APPLICANT: Heitler-Davis, Katharine
APPLICANT: Egan, Christina C.
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
ADDRESSEE: Laurence Weinberger
STREET: 882 S. Malack Street, Suite 103
CITY: West Chester
STATE: PA
COUNTRY: USA
ZIP: 19380-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,742
FILING DATE: 27-FEB-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Weinberger, Laurence
REGISTRATION NUMBER: 27,965
REFERENCE/DOCKET NUMBER: 3086-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 431-1703
TELEFAX: (610) 431-4181
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-09-032-742-8

Query Match 5.58: Score 91.5; DB 4; Length 471;
Best Local Similarity 20.4%; Pred. No. 0.1;

Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

45 LTIMLISNLFVLPLOSSTSH---ISFKDAFLGLIIP-----GY----- 85
Db 94 LVIMAVS-----LEKKLNATVIFLMSLAIDMLGFLVMPVSMILTYGYRMPDPSK 146
QY 86 -----YAALPNMOPSO-----VKDLTKCVAMLLTYCG 115
Db 147 LCAVWIVLDVLFSTASIMHLCALISLDRYVAIQNPDIHHSRFSRRTKAFKIIA-VMTISVG 205
QY 116 PVLDVLYHLNPK-----SILEDYHEFLNWSFRNFIFAPITEIFRTSMLLTYL 169
Db 206 ISMPIPVFGLODSKVEREGSCLAD--DNFVLIGSEVSF-FIPL-----TIMVITYFL 256
QY 170 NL-----IPHSOLSYOOLF-----OPSLFEGLAHAAHAY 199
Db 257 TIKSLOKEATLCVSDLTGRAKLASFSFLPOSSLSSEKLFORSIHREGSYTG----- 308
QY 200 EQLOEGSMTVSILTTCCFOILYTTLFGGLTKFVFRGGLMWC-----CIILHALC 251
Db 309 -----RRTMOSISNEOKAC-KVL-----GIVEFLFV-----VMCPFFITNIMAVICKESC 353
QY 252 N--IMGPPSPRLNLHFTVVDKRAGRISKLV-SIMMKCY 287
Db 354 NEDVIG---ALLNV-FWVIGYLSAVNPLVYTLFNKTY 387

RESULT 14
PCT-US93-00149-8

; Sequence 8, Application PC/TUS9300149
; GENERAL INFORMATION:
; APPLICANT: Weinshank, Richard L.
; APPLICANT: Branchek, Theresa
; APPLICANT: Hartlig, Paul R.
; TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
; NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00149
FILING DATE: 19930108
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39318
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
IMMEDIATE SOURCE:
CLONE: 5-HT2
PCT-US93-00149-8

Query Match 5.58: Score 91.5; DB 5; Length 471;

Best Local Similarity 20.4%; Pred. No. 0.1;

Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

45 LTIMLISNLFVLPLOSSTSH---ISFKDAFLGLIIP-----GY----- 85
Db 94 LVIMAVS-----LEKKLNATVIFLMSLAIDMLGFLVMPVSMILTYGYRMPDPSK 146
QY 86 -----YAALPNMOPSO-----VKDLTKCVAMLLTYCG 115
Db 147 LCAVWIVLDVLFSTASIMHLCALISLDRYVAIQNPDIHHSRFSRRTKAFKIIA-VMTISVG 205
QY 116 PVLDVLYHLNPK-----SILEDYHEFLNWSFRNFIFAPITEIFRTSMLLTYL 169
Db 206 ISMPIPVFGLODSKVEREGSCLAD--DNFVLIGSEVSF-FIPL-----TIMVITYFL 256
QY 170 NL-----IPHSOLSYOOLF-----OPSLFEGLAHAAHAY 199
Db 257 TIKSLOKEATLCVSDLTGRAKLASFSFLPOSSLSSEKLFORSIHREGSYTG----- 308
QY 200 EQLOEGSMTVSILTTCCFOILYTTLFGGLTKFVFRGGLMWC-----CIILHALC 251
Db 309 -----RRTMOSISNEOKAC-KVL-----GIVEFLFV-----VMCPFFITNIMAVICKESC 353
QY 252 N--IMGPPSPRLNLHFTVVDKRAGRISKLV-SIMMKCY 287
Db 354 NEDVIG---ALLNV-FWVIGYLSAVNPLVYTLFNKTY 387

RESULT 15

US-07-996-772A-11
; Sequence 11, Application US/07996772A
; Patent No. 5472866
; GENERAL INFORMATION:

THIS PAGE BLANK (JPT0)